



*Bienvenue!*

**ÉCOLE D'ÉTÉ FRANCOPHONE  
EN APPRENTISSAGE PROFOND**

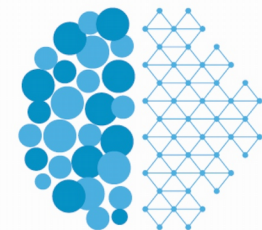
---

**21-25 août 2017**



**IVADO**

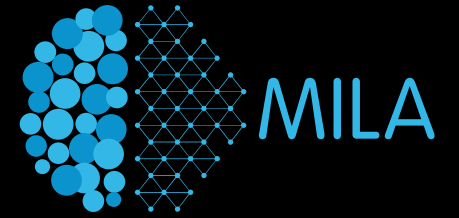
HEC Montréal  
Polytechnique Montréal  
Université de Montréal



**MILA**



Institut  
des algorithmes  
d'apprentissage  
de Montréal

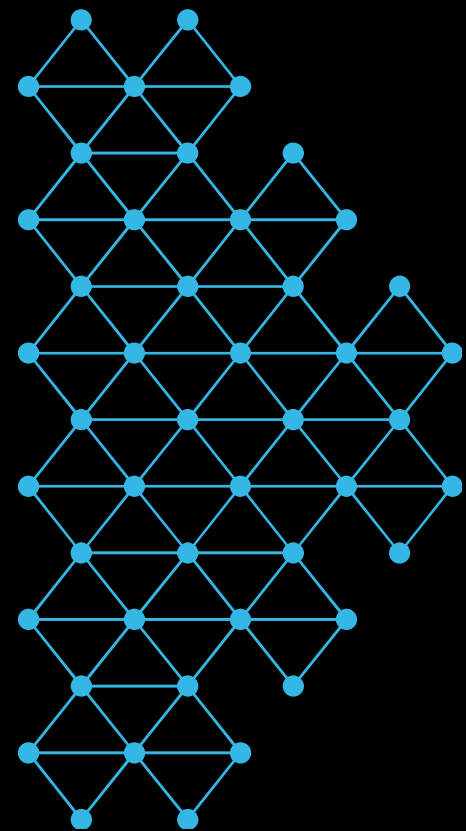


# Réseaux pour le langage

Gaétan Marceau Caron

École d'été francophone en apprentissage profond

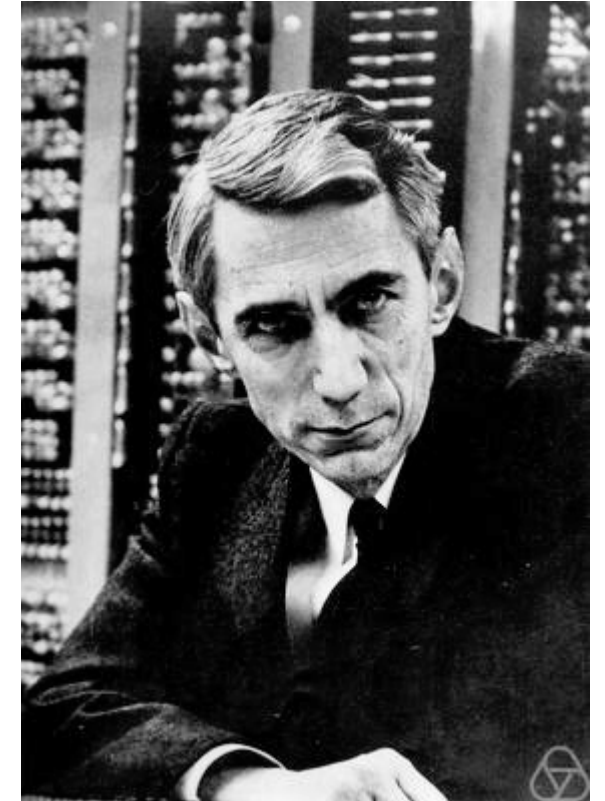
24 août 2017



# Les langues naturelles et les machines?

# Statistique de symboles

Approximation d'ordre 0 (Comptage de caractères)	XFOML RXKHRJFFJUJ ALPWXFWJXYJ FFJEYVJQCQSGHYD QPAAMKBZAACIBZLKJQD
Approximation d'ordre 1 (Comptage de 1-uple)	OCRO HLO RGWR NMIELWIS EU LL NBNSEBYA TH EEI ALHENHTTPA OOBTTVA NAH BRL
Approximation d'ordre 2 (Comptage de 2-uple)	ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILONASIVE TUCCOOWE AT TEASONARE FUSO TIZIN ANDY TOBE SEACE CTISBE
Approximation d'ordre 3 (Comptage de 3-uple)	IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID PONDENOME OF DEMONSTURES OF THE REPTAGIN IS REGOACTIONA OF CRE
Approximation d'ordre 1 (Comptage de 1-uple de mots)	REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME CAN DIFFERENT NATURAL HERE HE THE A IN CAME THE TO OF TO EXPERT GRAY COME TO FURNISHES THE LINE MESSAGE HAD BE THESE
Approximation d'ordre 2 (Comptage de 2-uple de mots)	THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED



Claude Shannon,  
*A mathematical theory  
of communication*  
(1948)

# Classification d'émotions

This is one of Crichton's best books. The characters of Karen Ross, Peter Elliot, Munro, and Amy are beautifully developed and their interactions are exciting, complex, and fast-paced throughout this impressive novel. And about 99.8 percent of that got lost in the film. Seriously, the screenplay AND the directing were horrendous and clearly done by people who could not fathom what was good about the novel. I can't fault the actors because frankly, they never had a chance to make this turkey live up to Crichton's original work. I know good novels, especially those with a science fiction edge, are hard to bring to the screen in a way that lives up to the original. But this may be the absolute worst disparity in quality between novel and screen adaptation ever. The book is really, really good. The movie is just dreadful.

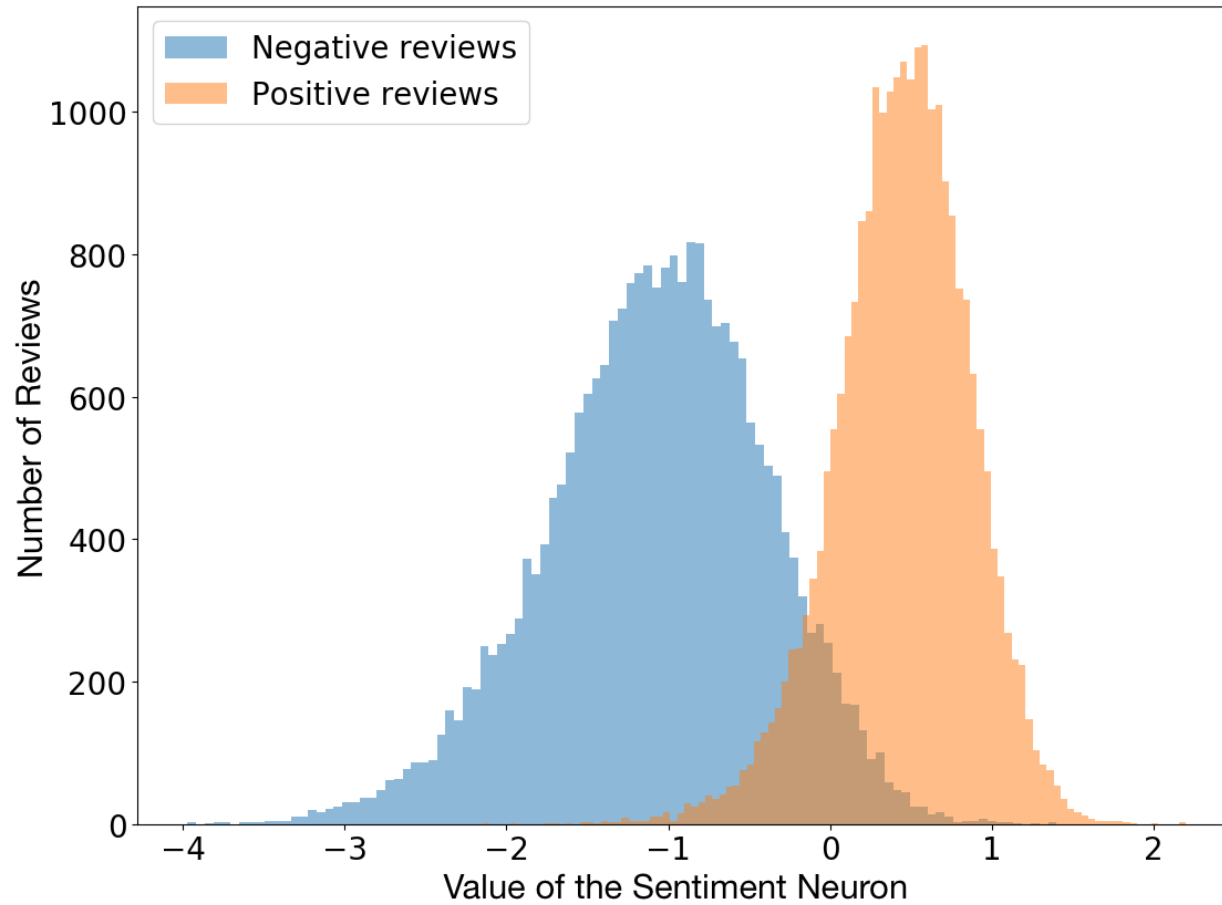
Classification au niveau du caractère

**Vert:** Positif

**Rouge:** Négatif

Radford, Alec, Rafal Jozefowicz, and Ilya Sutskever. "Learning to generate reviews and discovering sentiment." arXiv preprint arXiv:1704.01444 (2017).

# Classification d'émotions



Radford, Alec, Rafal Jozefowicz, and Ilya Sutskever. "Learning to generate reviews and discovering sentiment." arXiv preprint arXiv:1704.01444 (2017).

# Classification d'émotions

## Generating Reviews and Discovering Sentiment

---

### Sentiment fixed to positive

---

Just what I was looking for. Nice fitted pants, exactly matched seam to color contrast with other pants I own. Highly recommended and also very happy!

This product does what it is supposed to. I always keep three of these in my kitchen just in case ever I need a replacement cord.

### Sentiment fixed to negative

---

The package received was blank and has no barcode. A waste of time and money.

Great little item. Hard to put on the crib without some kind of embellishment. My guess is just like the screw kind of attachment I had.

Radford, Alec, Rafal Jozefowicz, and Ilya Sutskever. "Learning to generate reviews and discovering sentiment." arXiv preprint arXiv:1704.01444 (2017).

# Traduction automatique

Source	An admitting privilege is the right of a doctor to admit a patient to a hospital or a medical centre to carry out a diagnosis or a procedure, based on his status as a health care worker at a hospital.
Reference	Le privilège d'admission est le droit d'un médecin, en vertu de son statut de membre soignant d'un hôpital, d'admettre un patient dans un hôpital ou un centre médical afin d'y délivrer un diagnostic ou un traitement.
RNNenc-50	Un privilège d'admission est le droit d'un médecin de reconnaître un patient à l'hôpital ou un centre médical d'un diagnostic ou de prendre un diagnostic en fonction de son état de santé.
RNNsearch-50	Un privilège d'admission est le droit d'un médecin d'admettre un patient à un hôpital ou un centre médical pour effectuer un diagnostic ou une procédure, selon son statut de travailleur des soins de santé à l'hôpital.
Google Translate	Un privilège admettre est le droit d'un médecin d'admettre un patient dans un hôpital ou un centre médical pour effectuer un diagnostic ou une procédure, fondée sur sa situation en tant que travailleur de soins de santé dans un hôpital.

Bahdanau, D., Cho, K., & Bengio, Y. Neural machine translation by jointly learning to align and translate. ICLR (2015)



# De la recherche au grand public



**Mike Schroepfer**

23 h · 🌐 Directeur de la technologie de Facebook (4 août 2017)

A big part of bringing the world closer together is making sure that people can talk to each other. However, creating translations for more than 2 billion people on Facebook is no small task, especially when you consider the use of slang, colloquialisms and local customs.

Today we are announcing the latest milestone on our journey to bringing Facebook to everyone, regardless of what language they speak. All of our translations are now being powered by neural machine translation models, replacing our older, phrase based models. With these new models, we saw an average increase of 11% in accuracy and translations were delivered 2.5 times faster. Together these changes are helping people bridge language barriers to connect with others around the world. This switch builds on work from our FAIR team, and brings their industry leading research on using Convolutional Neural Nets (CNNs) for machine translation to production.

Here's a detailed post on how our Applied Machine Learning and Caffe2 teams made this a reality: [code.facebook.com/posts/289921871474277](https://code.facebook.com/posts/289921871474277)

Une grande partie du rapprochement du monde est de faire en sorte que les gens puissent se parler. Cependant, créer des traductions pour plus de 2 milliards de personnes sur facebook n'est pas une tâche de petite taille, surtout lorsque vous considérez l'utilisation d'argot, de expressions et de coutumes locales.


Aujourd'hui, nous annonçons la dernière étape de notre voyage pour amener facebook à tout le monde, quelle que soit la langue qu'ils parlent. Toutes nos traductions sont maintenant alimentées par des modèles de traduction de machine neural, en remplacement de nos anciens modèles de phrases. Avec ces nouveaux modèles, nous avons vu une augmentation moyenne de 11 % en précision et des traductions ont été livrées 2.5 fois plus vite. Ensemble, ces changements aident les gens à surmonter les barrières linguistiques pour se connecter avec d'autres personnes dans le monde. Ce commutateur s'appuie sur le travail de notre équipe, et apporte à leur industrie des recherches de pointe sur l'utilisation de convolutional neural (cns) pour la traduction automatique à la production.

Voici un article détaillé sur la façon dont nos équipes d'apprentissage de machines appliquées et les équipes caf2 ont fait de cette situation une réalité : [code.facebook.com/posts/289921871474277](https://code.facebook.com/posts/289921871474277)

⚙️ · Notez cette traduction

# Question-réponse - Chatbot

- Watson gagne à Jeopardy! (2011)
- Erreur de parcours du chatbot Tay (2016)
- Sujet de recherche très actif!



The image shows a screenshot of the MILA website. At the top, there is a dark navigation bar with the MILA logo on the left and the words 'ACCUEIL', 'MILA', and 'RECHERCHE' on the right. Below the navigation bar, there is a light gray section containing a news article snippet. The snippet starts with the date '20 JUIL 2017' and the title 'LE DÉFI D'INTELLIGENCE CONVERSATIONNELLE NIPS CHERCHE DES ÉVALUATEURS HUMAINS'. The main text of the snippet reads: 'Joignez-vous à la compétition ConvAI comme évaluateur humain pour créer une nouvelle génération de robot-clavardeurs vraiment intelligents'.



IBM Watson  
(2011)



Microsoft  
Chatbot Tay  
(2016)

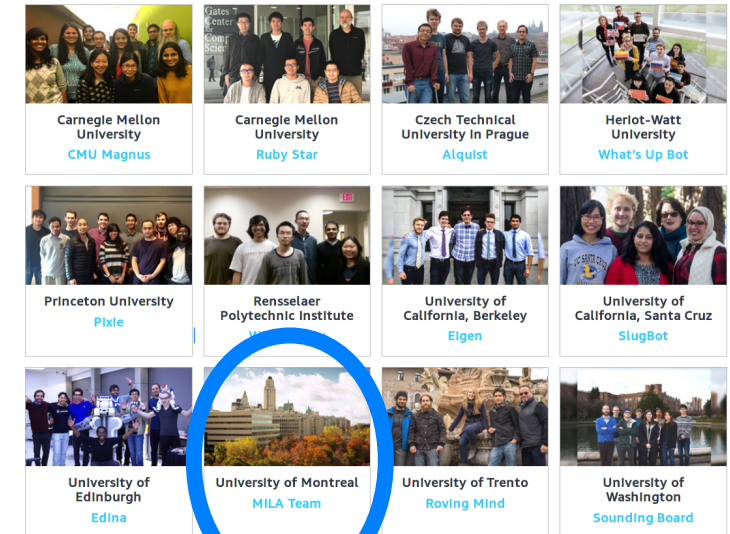
# Chatbot Alexa

## What is the goal of the challenge?

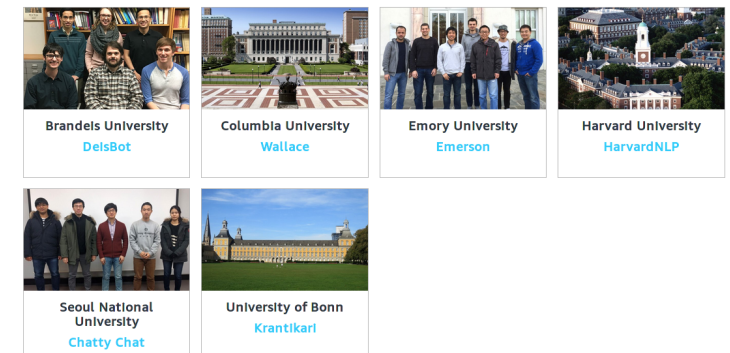
The goal of the challenge is to advance several areas of conversational AI including knowledge acquisition, natural language understanding (NLU), natural language generation (NLG), context modeling, commonsense reasoning and dialog planning. The grand challenge for the 2017 Alexa Prize is to create a socialbot that converses coherently and engagingly with humans on popular topics for 20 minutes.

## Meet the 2016-2017 Teams

### Sponsored Teams



### Unsponsored Teams



# Impact du DL sur le NLP

- Traduction automatique
- Modélisation du langage
- Étiquetage morpho-syntaxique (Part-Of-Speech Tagging)
- Chunking
- Reconnaissance des entités nommées (Named Entity Recognition)
- Reconnaissance du rôle sémantique (Semantic Role Labeling)
- Analyse syntaxique (Parsing)
- Reconnaissance de sentiments (Sentiment analysis)
- Reformulation automatique (Paraphrasing)
- Question-réponse (Q&A)
- Désambiguïsation lexicale (Word-Sense disambiguation)
- Résumé automatique (Automatic summarization)
- Génération
- ...

# Représentation des mots

- Hypothèse: les mots sont des unités atomiques

Un vecteur binaire par mot (one-hot vector)

chat = [0 0 **1** 0 0 0 0 0 0 0 0 0 0 0 ... 0]

- Similarité des mots:

chat = [0 0 **1** 0 0 0 0 0 0 0 0 0 0 0 0 ... 0]

chien = [0 0 0 0 0 0 0 0 0 0 **1** 0 0 0 0 ... 0]

maison = [0 0 0 0 0 0 0 0 0 0 0 0 0 **1** 0 ... 0]

- Tous les mots différents sont à une distance de 2 !

# Représentation des phrases

- Bag of words:

Le chat est dans la maison avec le chien:

[0 0 **1** 0 0 0 0 0 0 0 **1** 0 0 **1** 0...0]

chat = [0 0 **1** 0 0 0 0 0 0 0 0 0 0 0 0...0]

chien = [0 0 0 0 0 0 0 0 0 0 **1** 0 0 0 0...0]

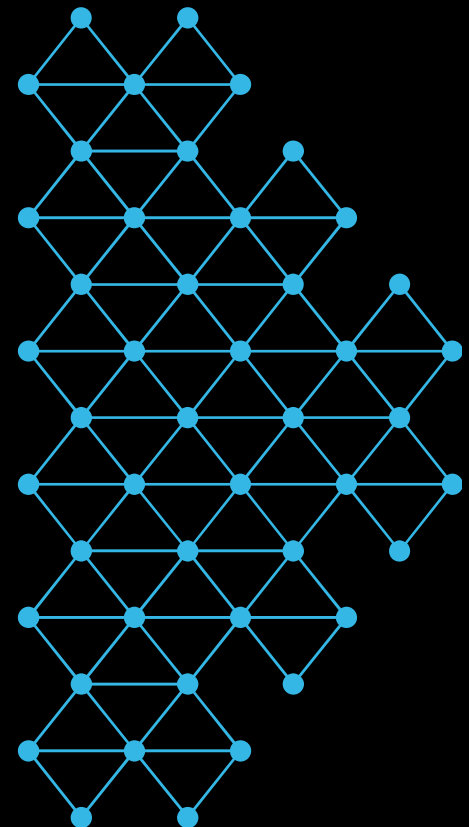
maison = [0 0 0 0 0 0 0 0 0 0 0 0 0 **1** 0...0]

- On perd l'ordre des mots

# Hypothèse distributionnelle

“You shall know a word by the company it keeps”

Firth, J. R. 1957:11



# Exemple

Je suis heureux de voir qu'il y a peut-être une petite lueur **d'intelligence** et de clairvoyance. Le langage et **l'intelligence** nous ont été donnés pour nous entendre et non pour nous battre entre nous. Dans l'économie de la connaissance, **l'intelligence** économique et technologique constitue un élément essentiel pour bâtir. Région pauvre en matières premières, le Land investit dans son capital le plus précieux, **l'intelligence**. Aucune force sensible ne peut commander à **l'intelligence** et à la volonté et les rejoindre dans leurs actes. Le savoir n'est jamais seulement l'œuvre de **l'intelligence**. Il s'agit d'appliquer les règles avec **intelligence** et autorité pour favoriser la croissance et l'emploi. Il y a aussi **l'intelligence** intuitive, car l'information empirique que nous possédons nous enseigne que nous n'agissons pas correctement. Ce serait preuve d'arrogance que d'affirmer que **l'intelligence** de l'homme peut parer à tous les risques engendrés par leur seule existence. Ses recherches présentent huit formes **d'intelligence** qui peuvent être développées et utilisées à travers l'enseignement des arts. Il identifie huit sortes **d'intelligence** que l'on retrouve à des degrés divers chez chaque individu. L'apprentissage coopératif, l'enseignement à niveaux multiples, **l'intelligence** multiple, les styles d'apprentissage et d'enseignement Des traducteurs potentiels subissent des altérations neurologique en symbiose avec des **intelligences** extra terrestres.

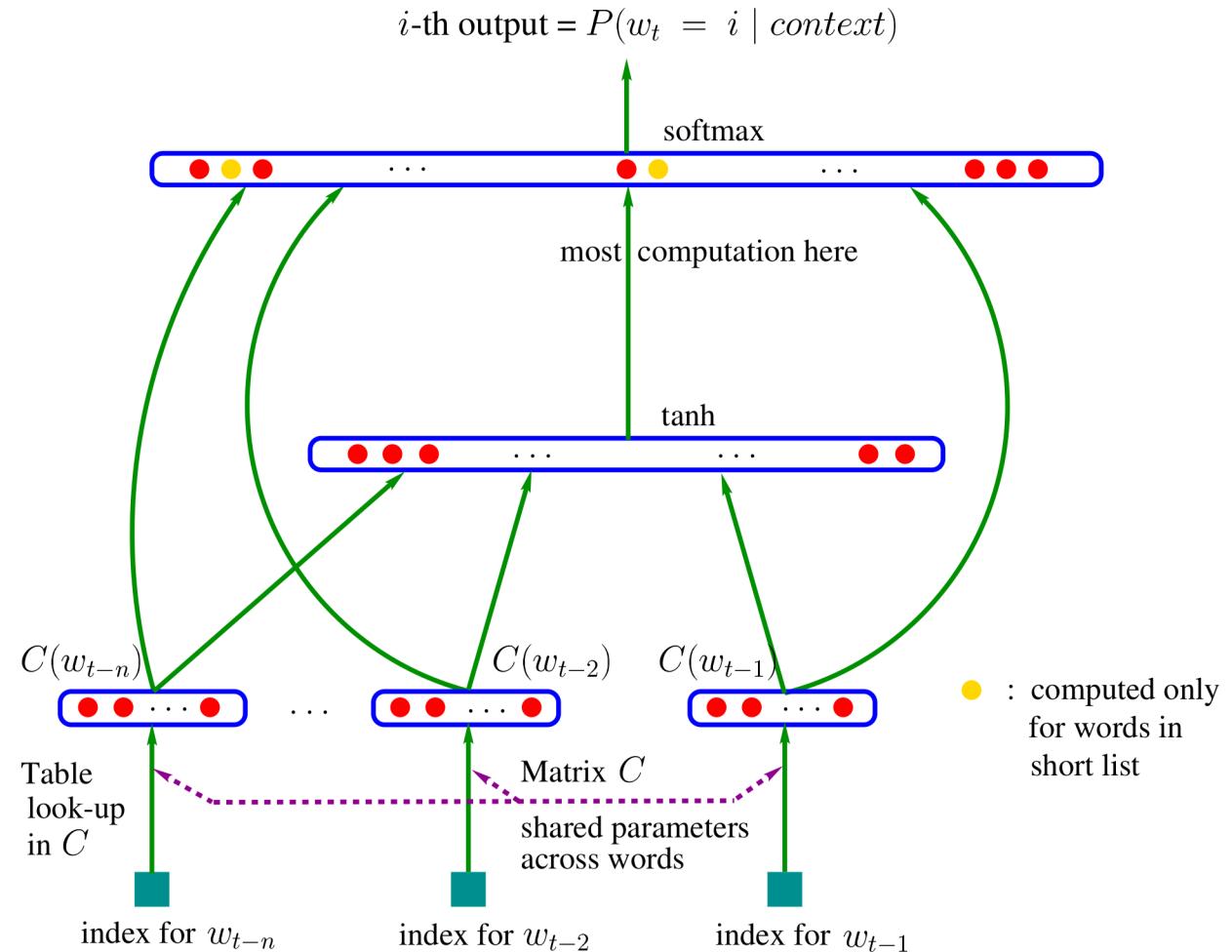
Sources: Wikipedia

<http://www.linguee.fr/francais-anglais/search?source=auto&query=intelligence>



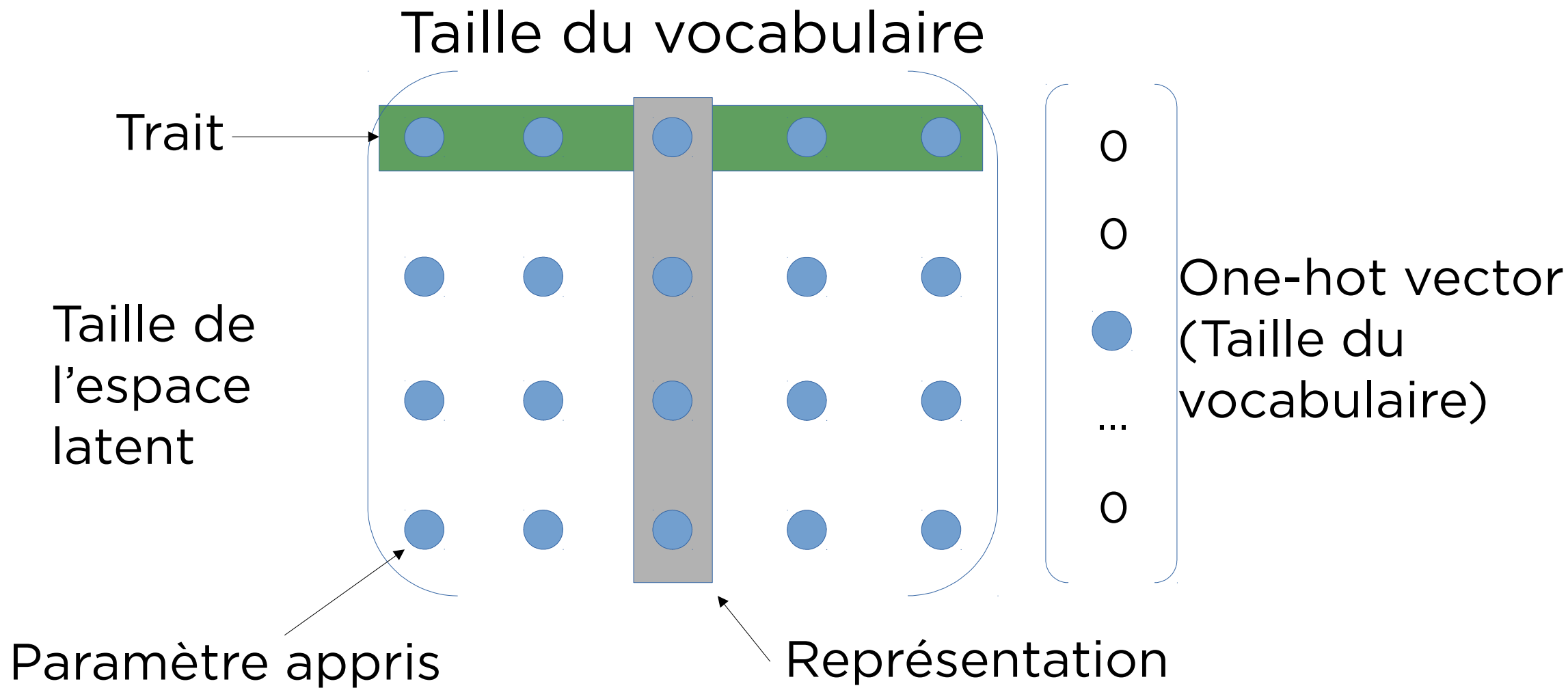
# Représentation distribuée

- Apprendre un **espace latent** dans lequel on capture les similarités et les relations entre les mots
- Encoder la sémantique à partir du contexte du mot courant
- Représentation distribuée permet de généraliser

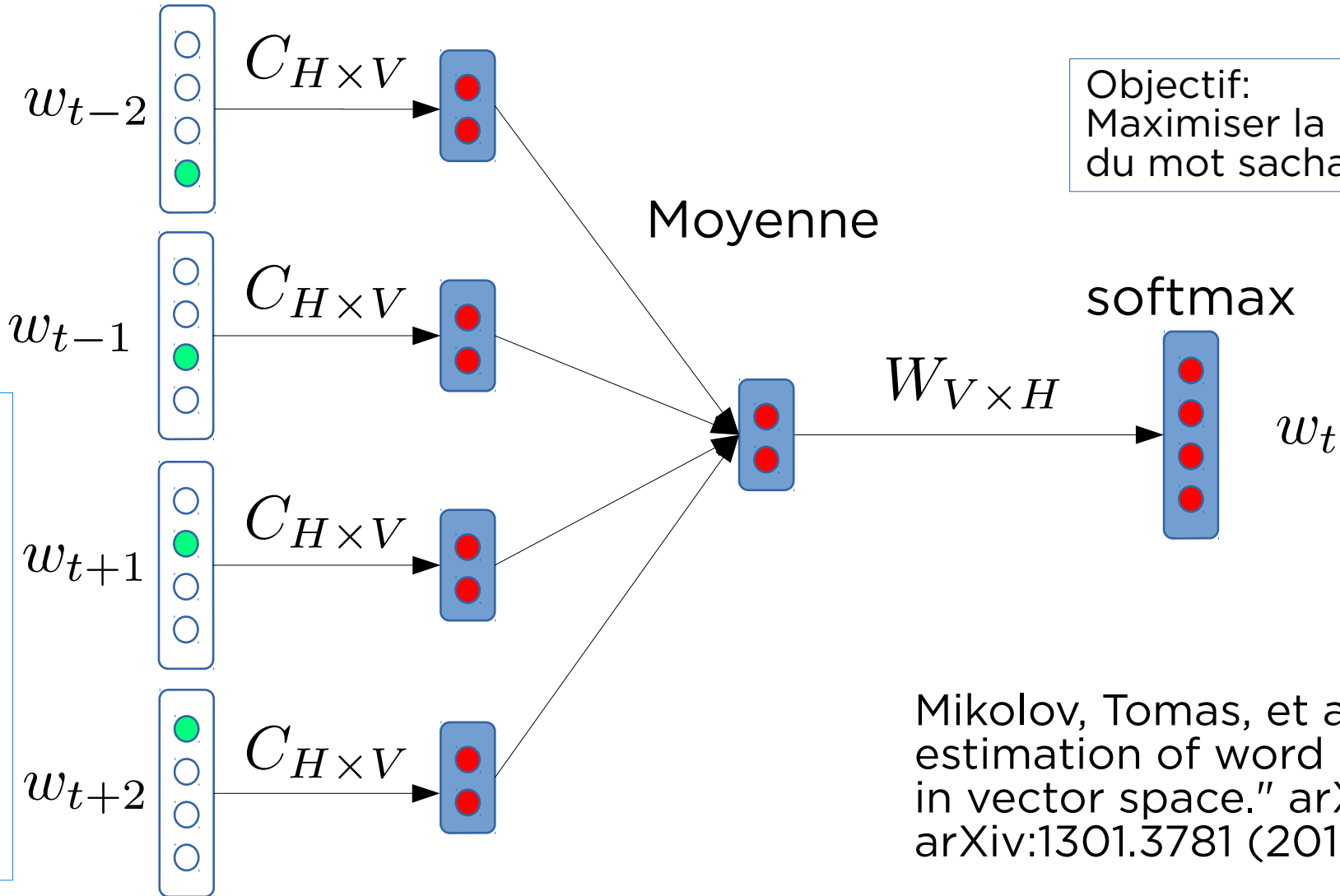


Bengio, Yoshua, et al. "A neural probabilistic language model." Journal of machine learning research 3.Feb (2003): 1137-1155.

# La transformation (Matrice) C



# Continuous Bag-of-Words Model



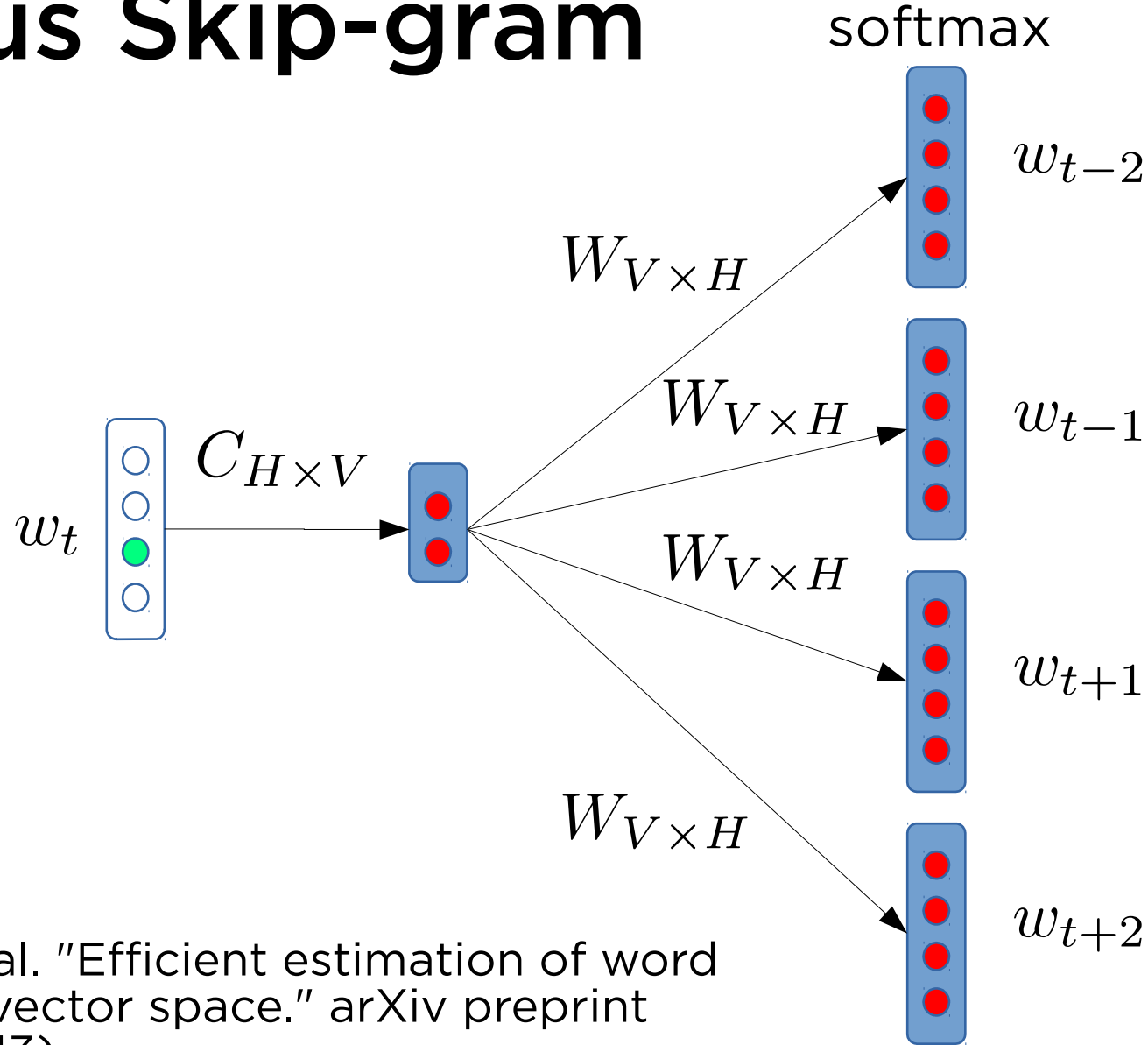
Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space." arXiv preprint arXiv:1301.3781 (2013).

# Continuous Skip-gram Model

Dataset (milliard):  
(mot, contexte)

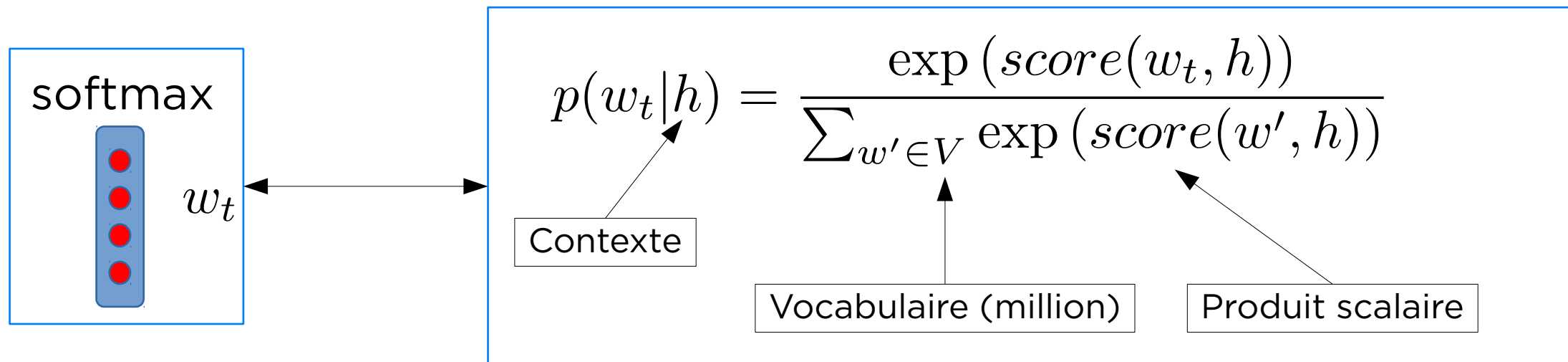
H: nombre  
de neurones de  
la couche cachée  
(millier)

V: taille du  
Vocabulaire  
(million)



Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space." arXiv preprint arXiv:1301.3781 (2013).

# Noise-Contrastive Training



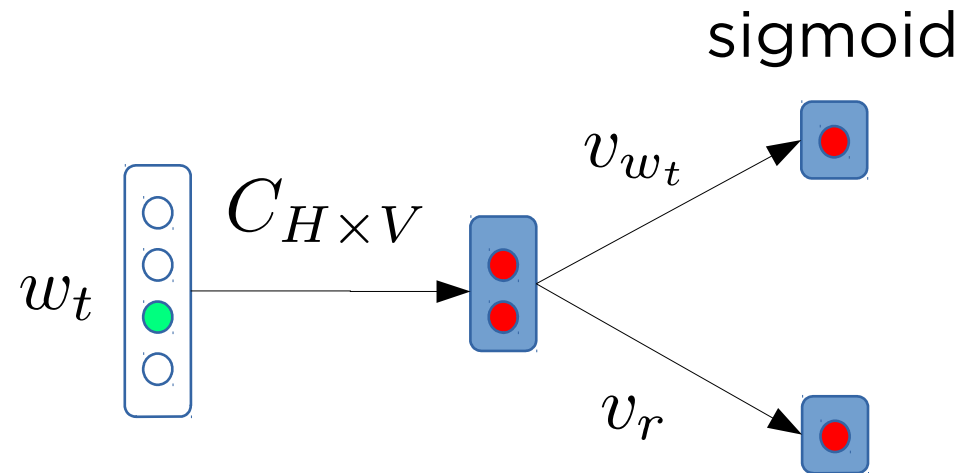
Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space." arXiv preprint arXiv:1301.3781 (2013).

# Noise-Contrastive Training

Dataset (milliard):  
(mot, contexte)

H: nombre  
de neurones de  
la couche cachée  
(millier)

V: taille du  
Vocabulaire  
(million)



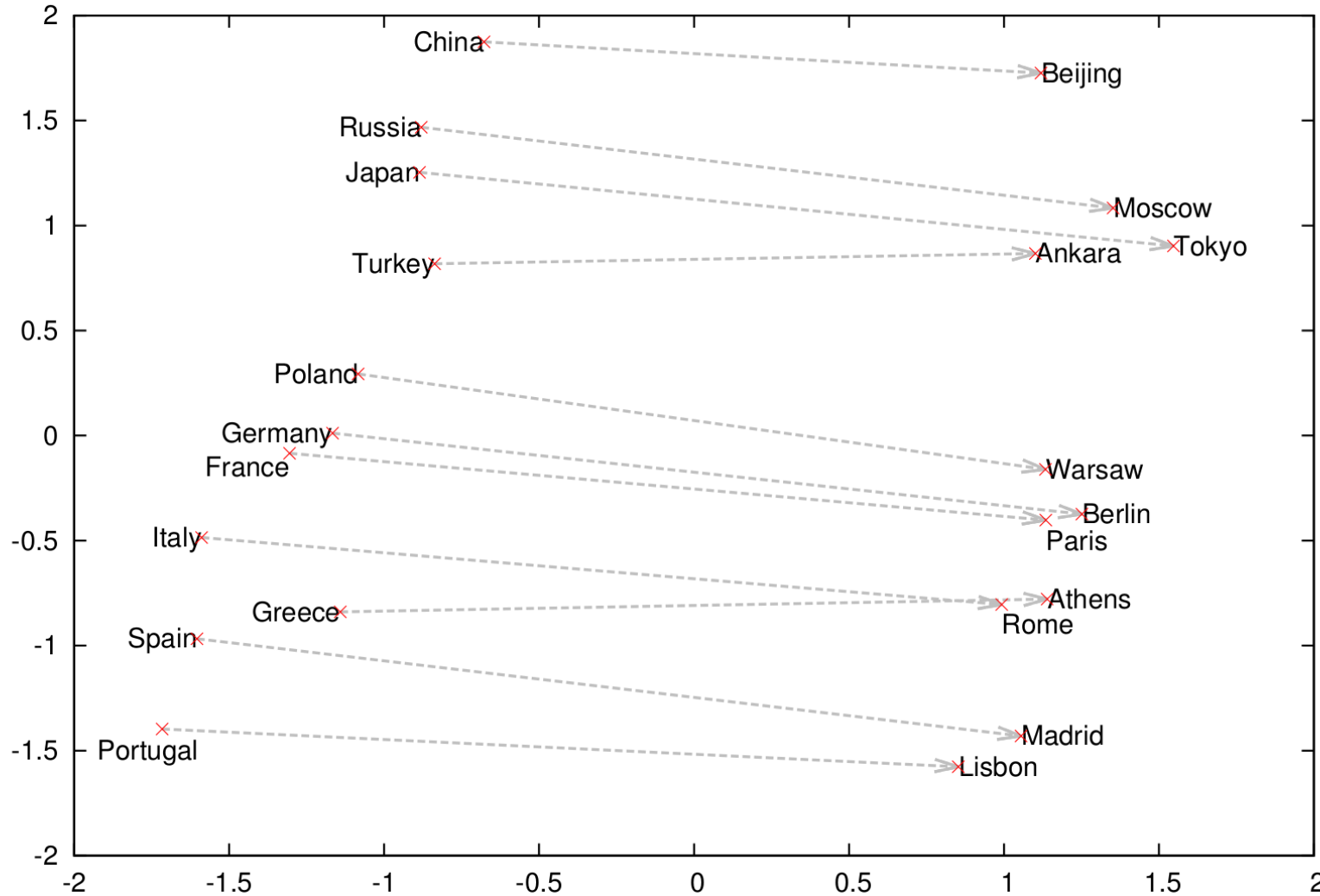
Objectif:  
Prédire si (mot, contexte) est  
dans la base d'apprentissage

On tire le contexte aléatoirement, e.g.,  
distribution empirique des mots

Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space." arXiv preprint arXiv:1301.3781 (2013).

# Évaluer la représentation distribuée

Country and Capital Vectors Projected by PCA



Mikolov, Tomas, et al. "Distributed representations of words and phrases and their compositionality." Advances in neural information processing systems. 2013.

# Évaluer la représentation distribuée

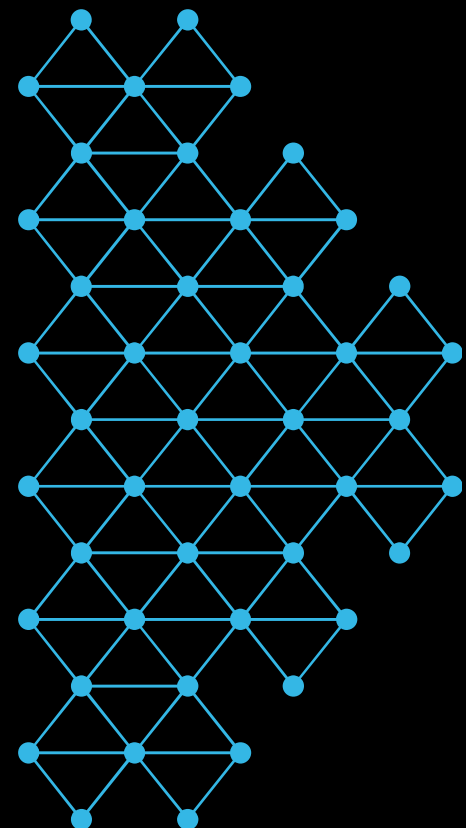
Table 8: *Examples of the word pair relationships, using the best word vectors from Table 4 (Skip-gram model trained on 783M words with 300 dimensionality).*

Relationship	Example 1	Example 2	Example 3
France - Paris	Italy: Rome	Japan: Tokyo	Florida: Tallahassee
big - bigger	small: larger	cold: colder	quick: quicker
Miami - Florida	Baltimore: Maryland	Dallas: Texas	Kona: Hawaii
Einstein - scientist	Messi: midfielder	Mozart: violinist	Picasso: painter
Sarkozy - France	Berlusconi: Italy	Merkel: Germany	Koizumi: Japan
copper - Cu	zinc: Zn	gold: Au	uranium: plutonium
Berlusconi - Silvio	Sarkozy: Nicolas	Putin: Medvedev	Obama: Barack
Microsoft - Windows	Google: Android	IBM: Linux	Apple: iPhone
Microsoft - Ballmer	Google: Yahoo	IBM: McNealy	Apple: Jobs
Japan - sushi	Germany: bratwurst	France: tapas	USA: pizza

Sun: McNealy

Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space." arXiv preprint arXiv:1301.3781 (2013).





# Hypothèse de Compositionnalité

En mathématiques, sémantique, et philosophie du langage, le principe de compositionnalité est le principe selon lequel la signification d'une expression complexe est définie par les significations des expressions la composant, et par les règles employées pour les combiner. - Wikipedia

# Cadre de travail

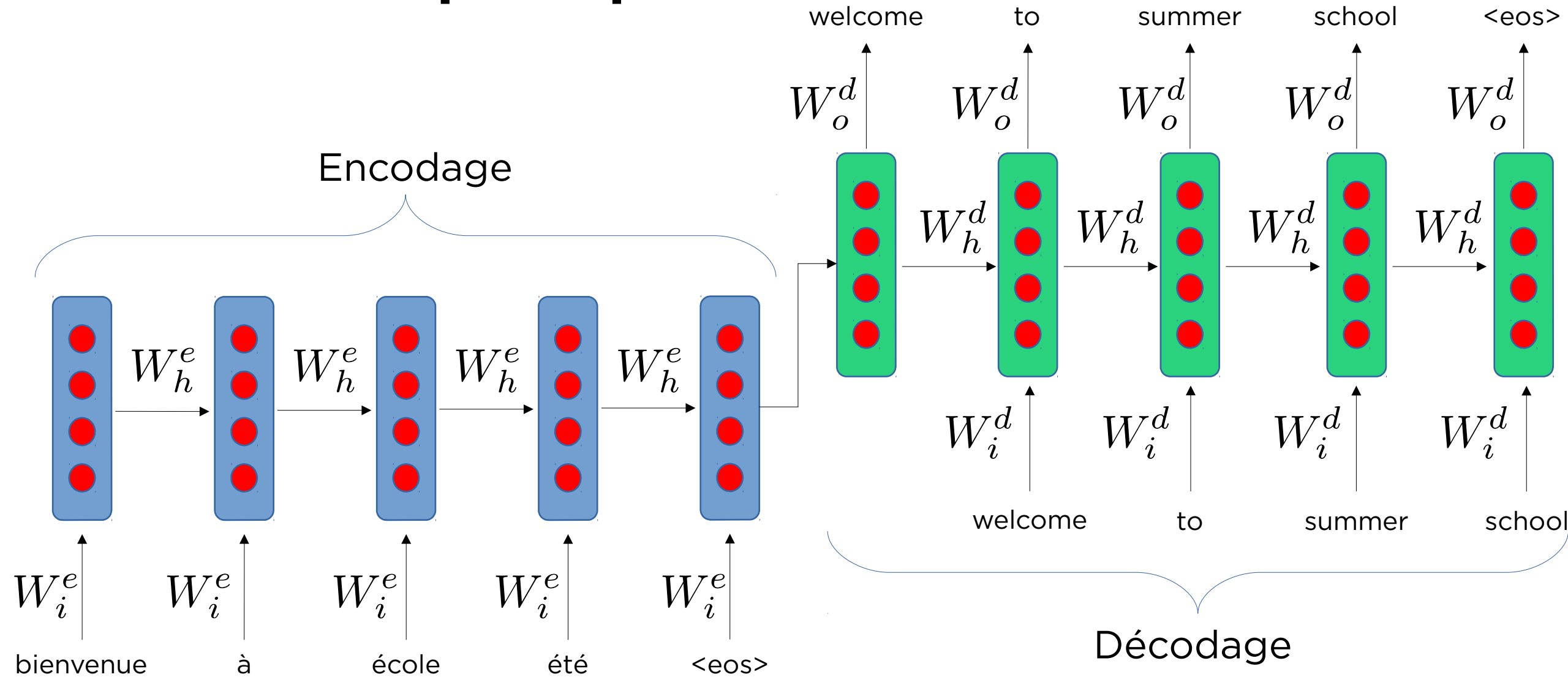
- Modélisation d'une **séquence cible** à partir d'une **séquence source**
- Données d'entraînement:  
couple de séquences **(source, cible)**
- Fonction de coût:  
maximiser la probabilité des données
- Applications spécifiques:  
**Traduction**, Q&A, résumé automatique, etc.

# Modèle seq2seq

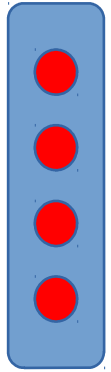
Kyunghyun Cho, Bart Merriënboer, Caglar Gulcehre, Fethi Bougares, Holger Schwenk, et al.. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. EMNLP, 2014, Unknown, Unknown or Invalid Region. 2014.

Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." Advances in neural information processing systems. 2014.

# Modèle seq2seq

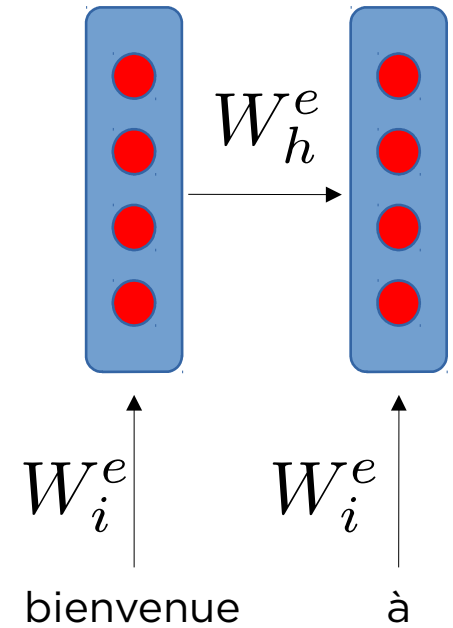


# Modèle seq2seq

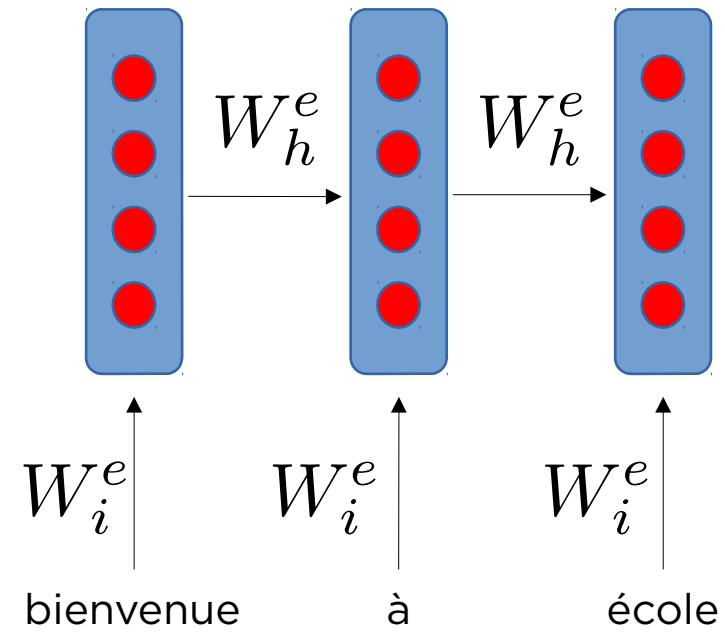


$W_i^e$   
↑  
bienvenue

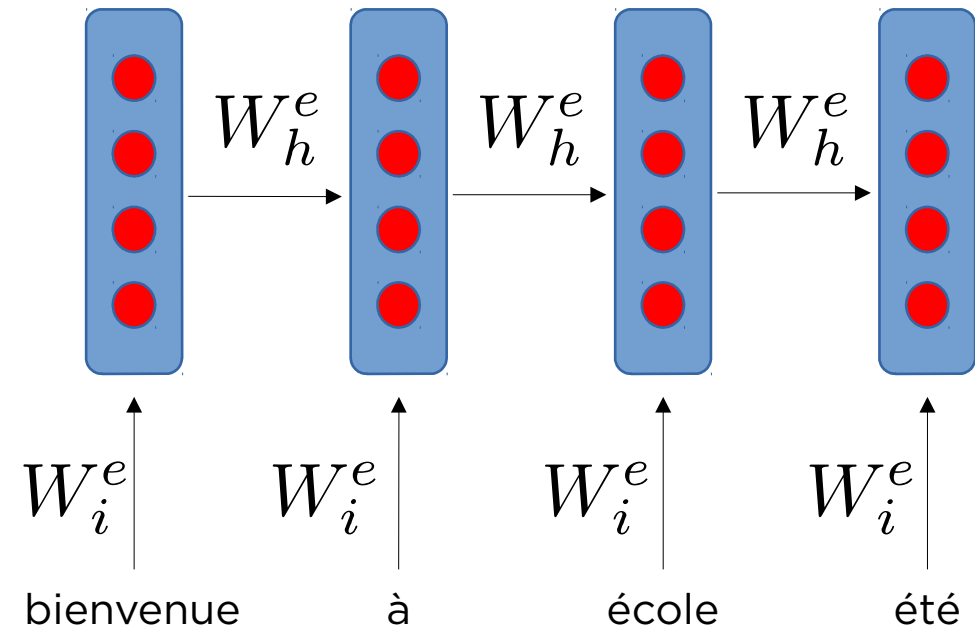
# Modèle seq2seq



# Modèle seq2seq

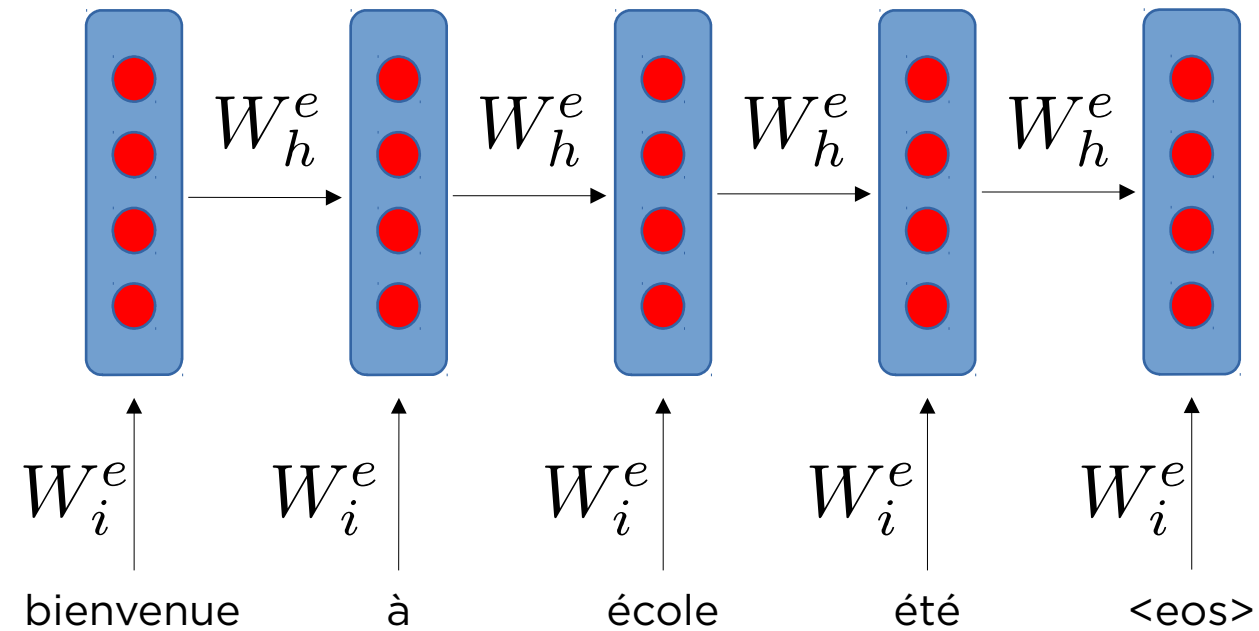


# Modèle seq2seq

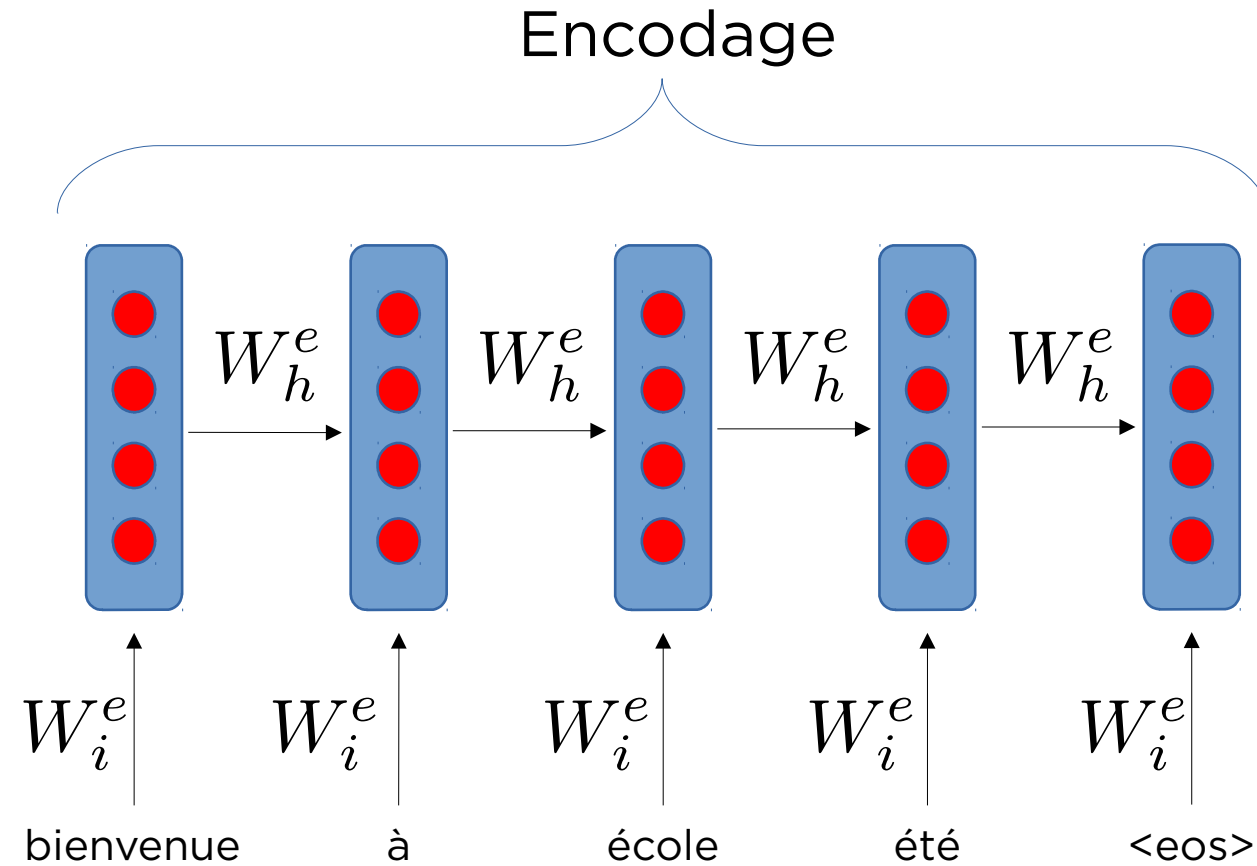




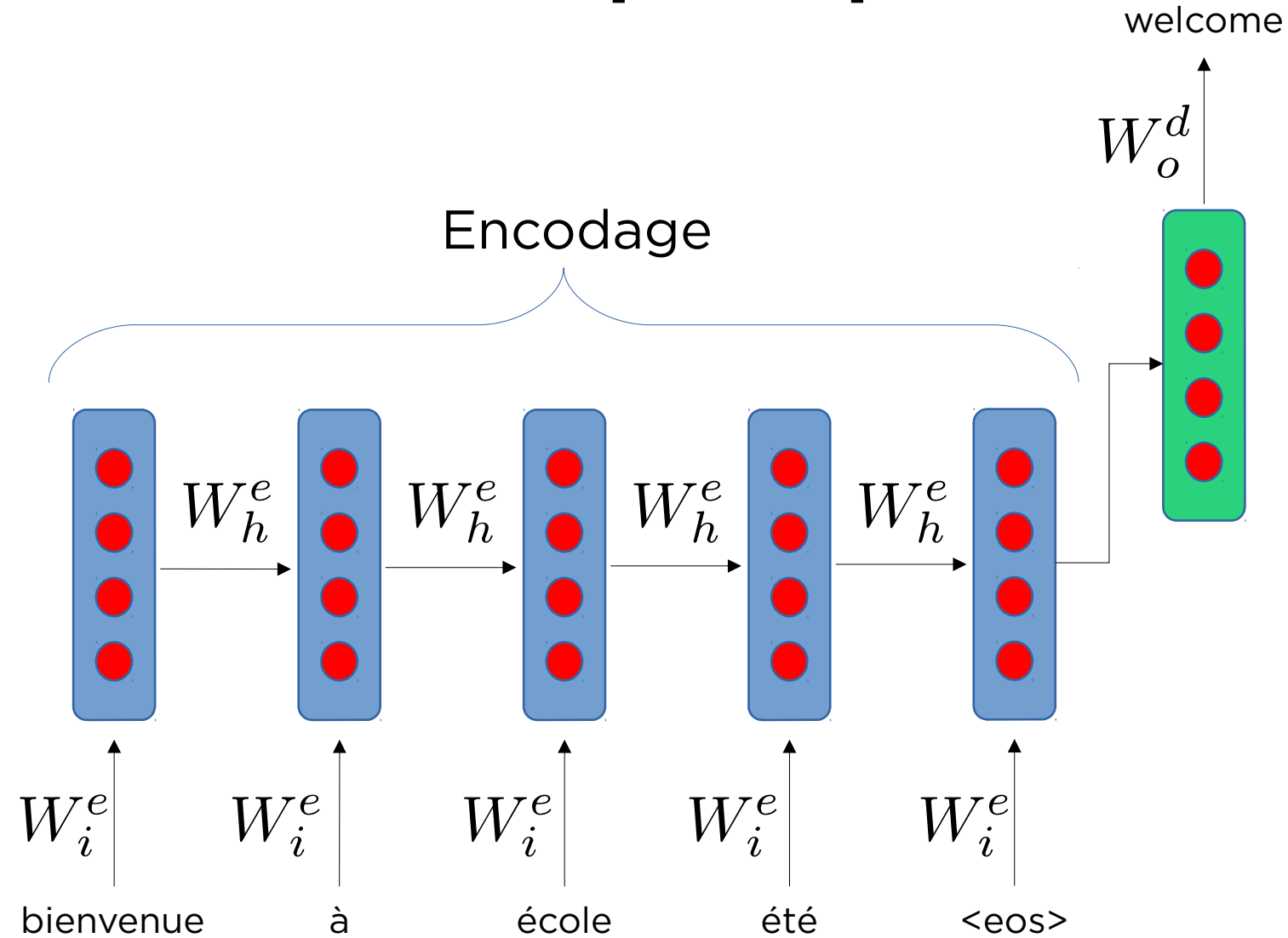
# Modèle seq2seq



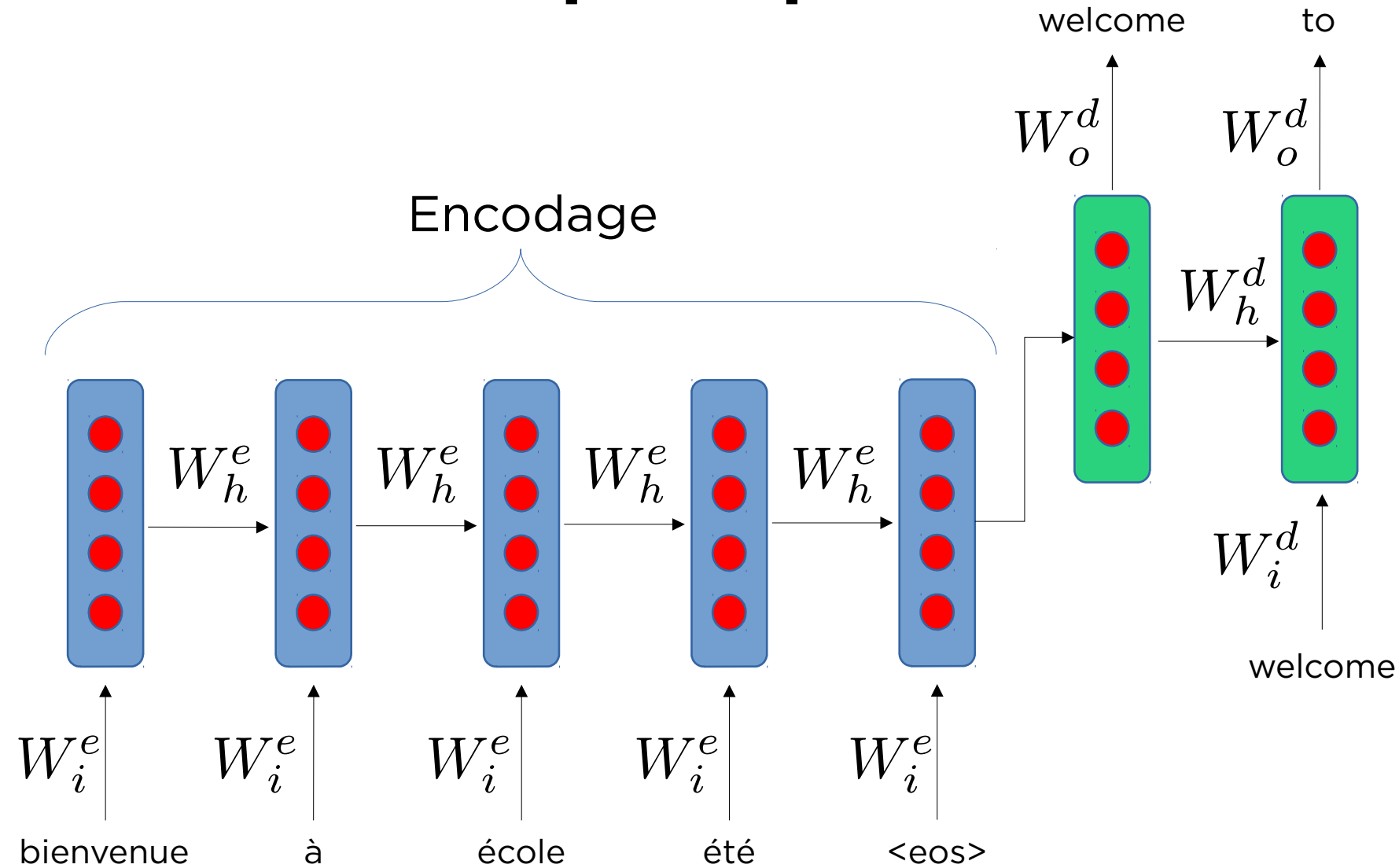
# Modèle seq2seq



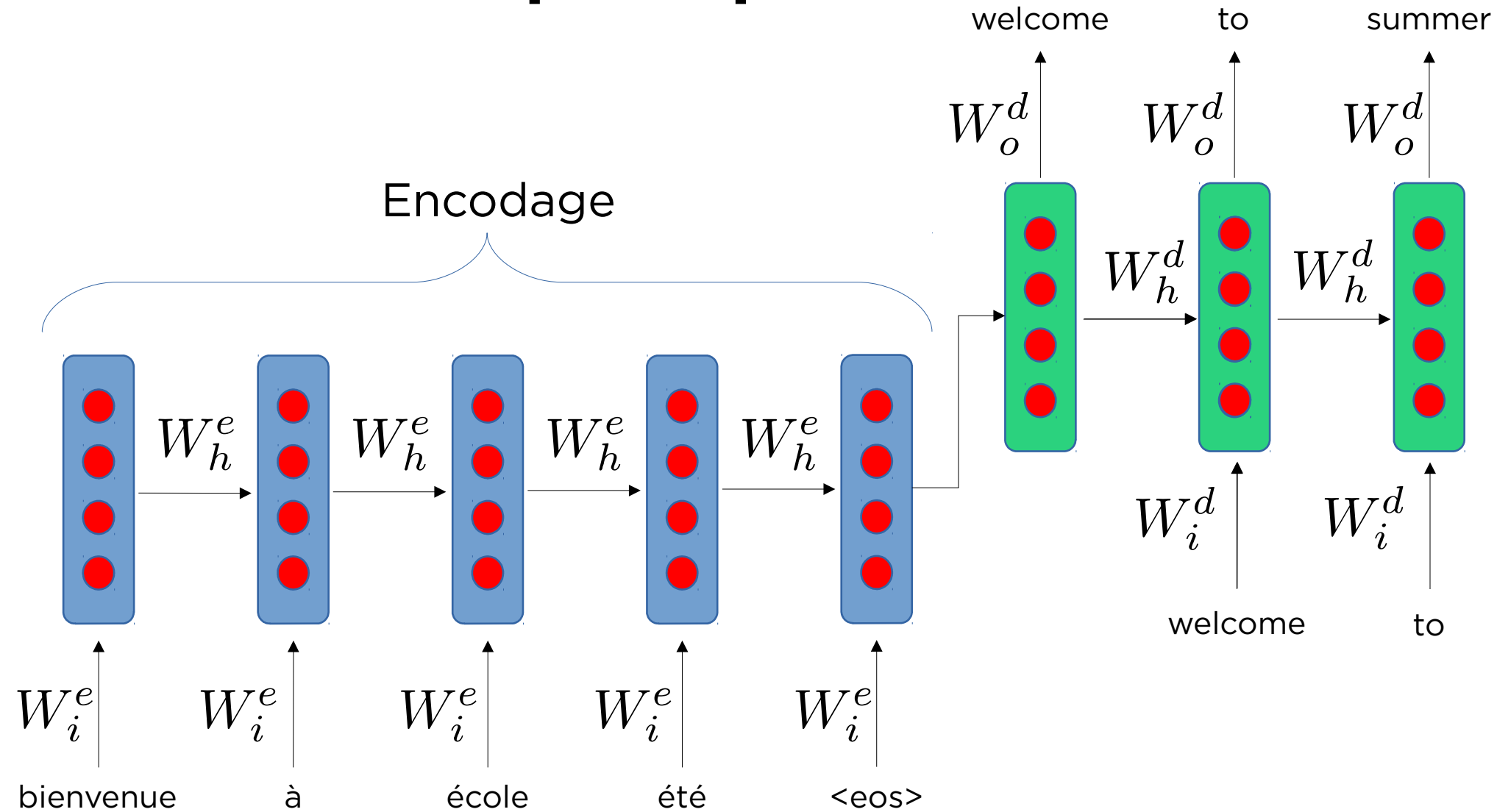
# Modèle seq2seq



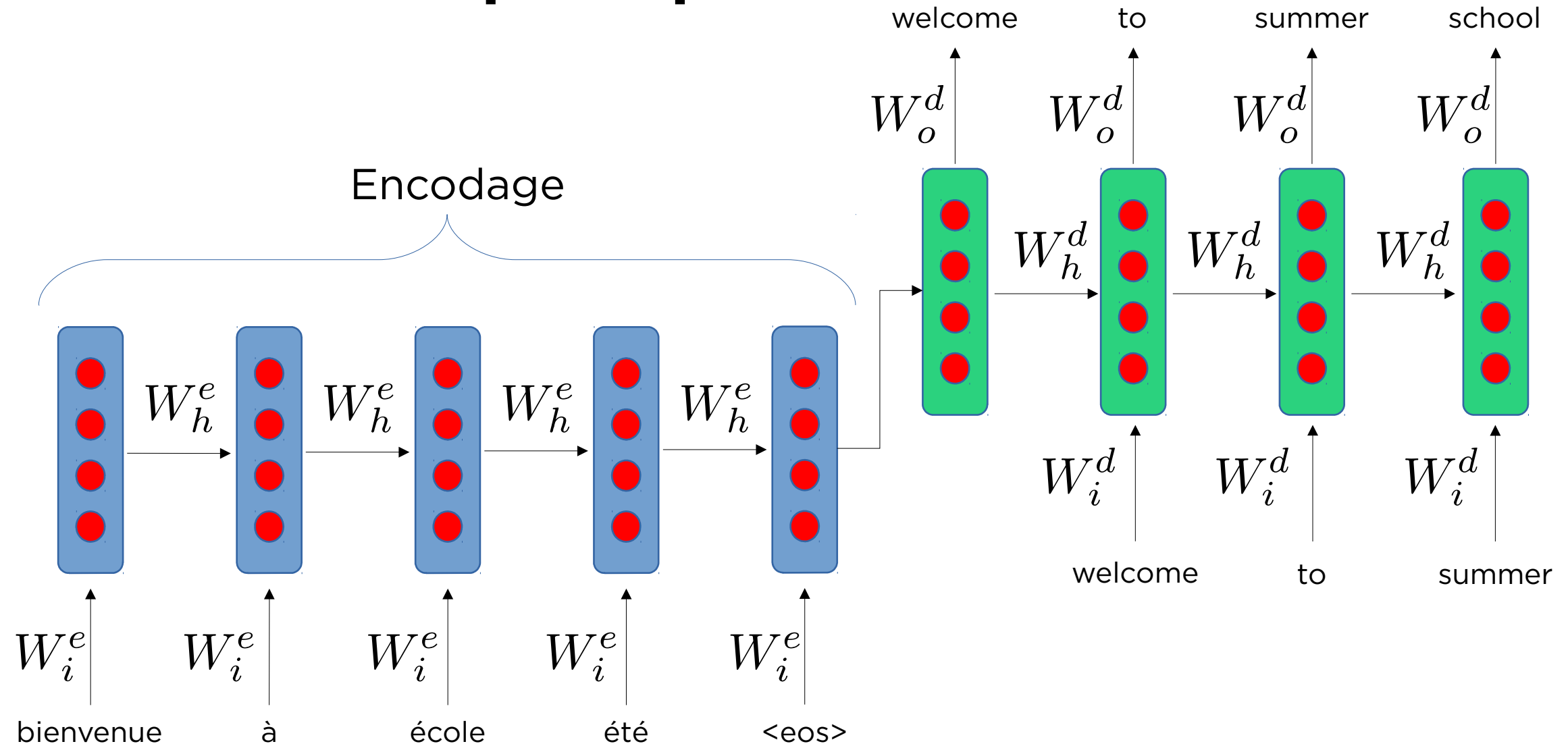
# Modèle seq2seq



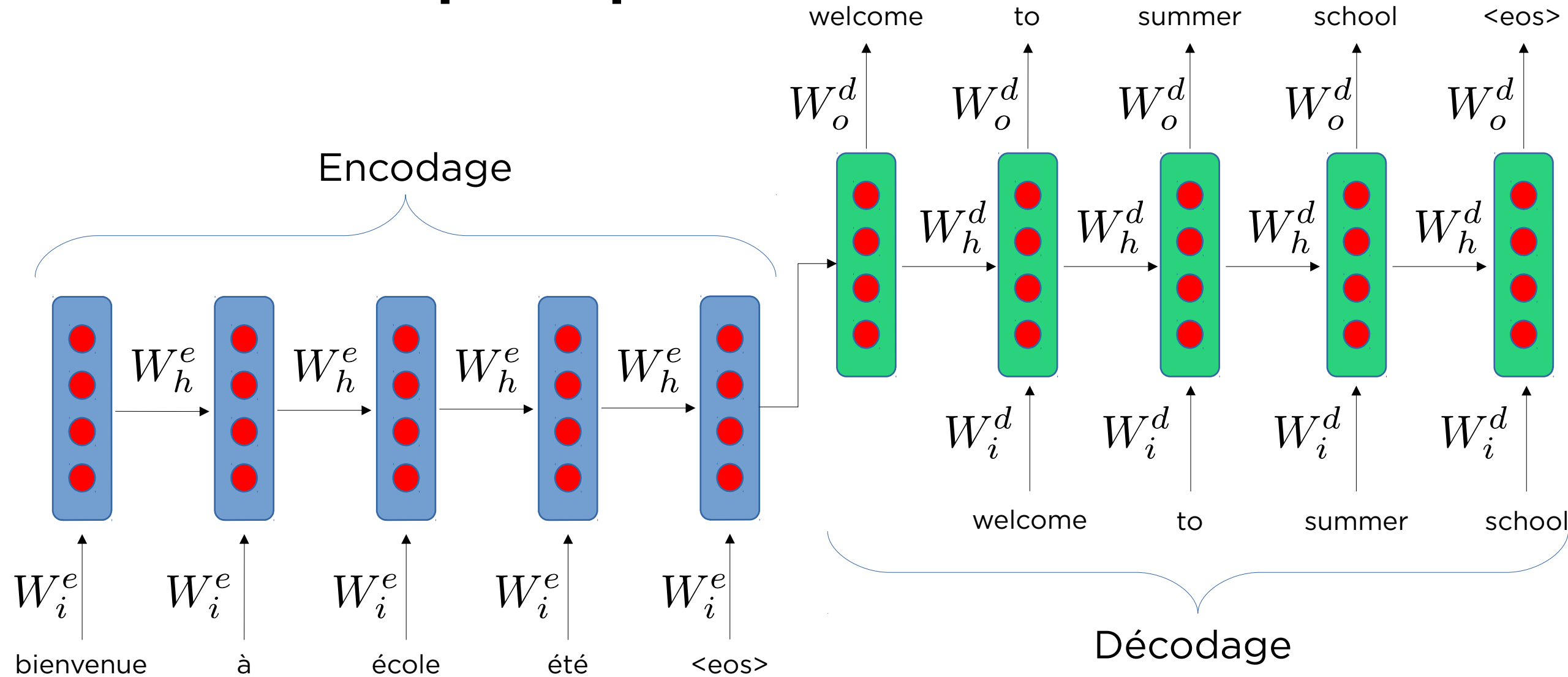
# Modèle seq2seq



# Modèle seq2seq

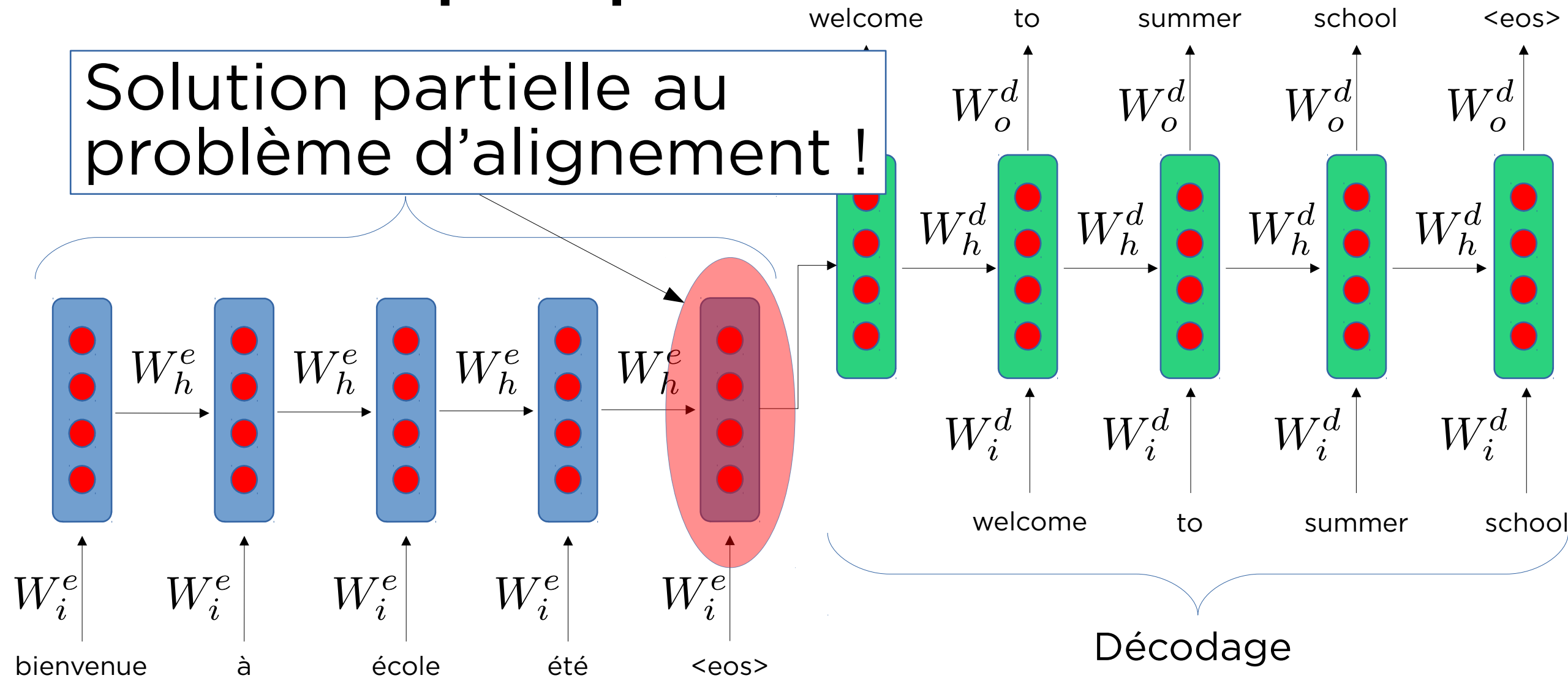


# Modèle seq2seq



# Modèle seq2seq

Solution partielle au problème d'alignement !





# Le problème de l'alignement

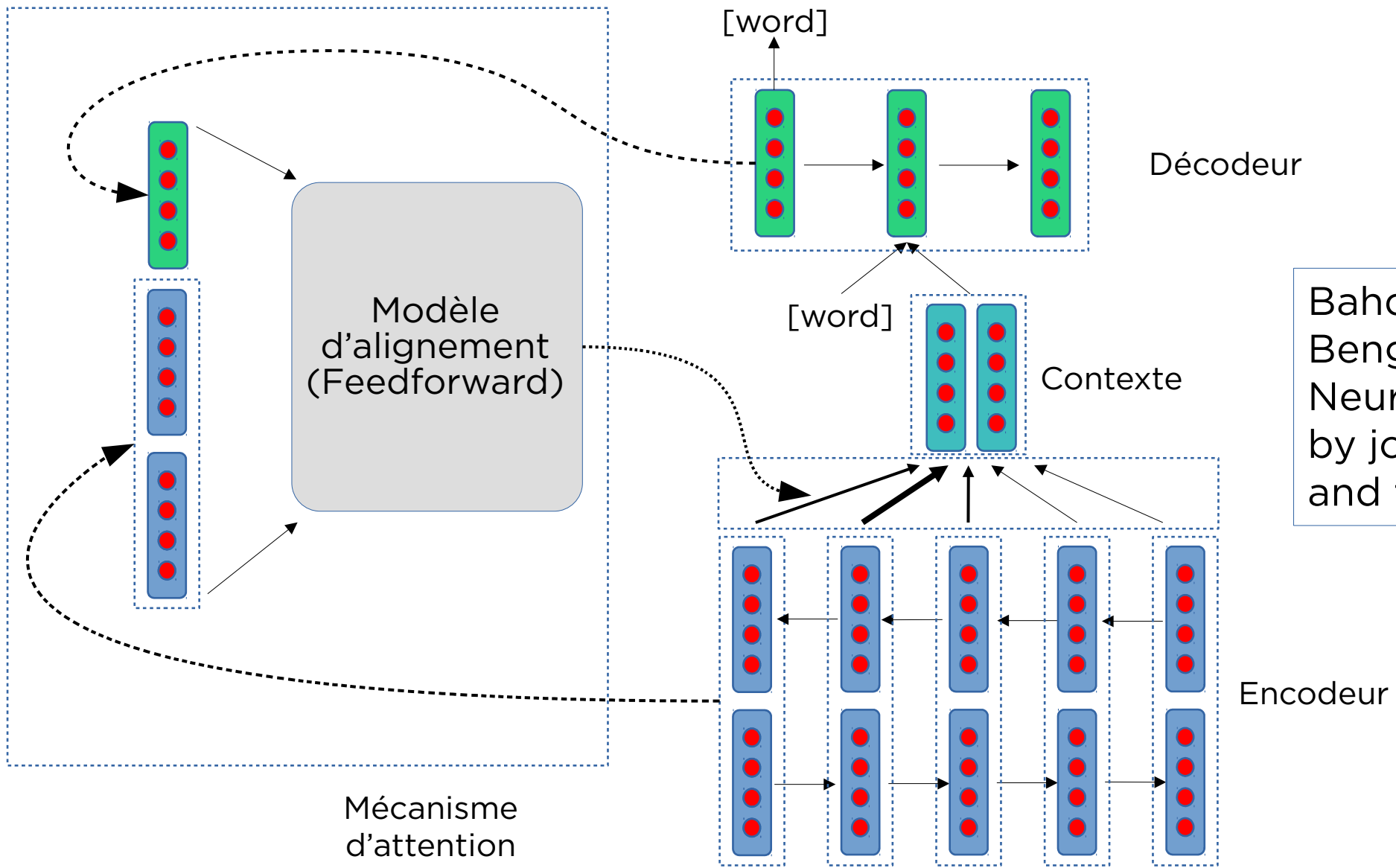
- Les ordres de la séquence source et la séquence cible ne sont pas forcément les mêmes
- Seq2seq semble être très générique
- Aucune issue...
- Deus ex Machina ?

# Le problème de l'alignement

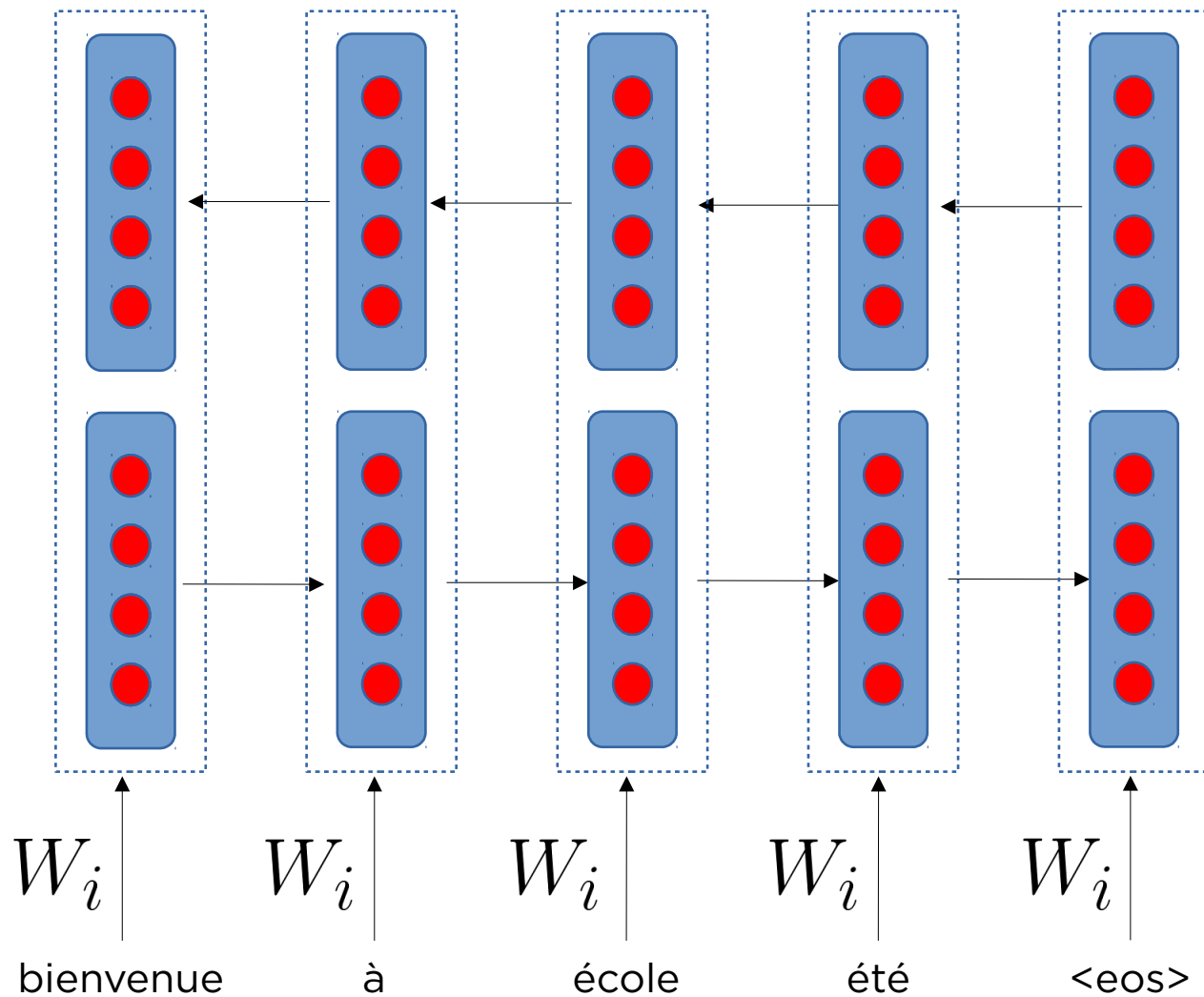
- L'ordre entre la séquence source et la séquence cible n'est pas forcément respecté
- Seq2seq sans attention : approche générique
- Attention : mécanisme d'alignement
- Deus ex Machina ?

**Le mécanisme d'attention !**

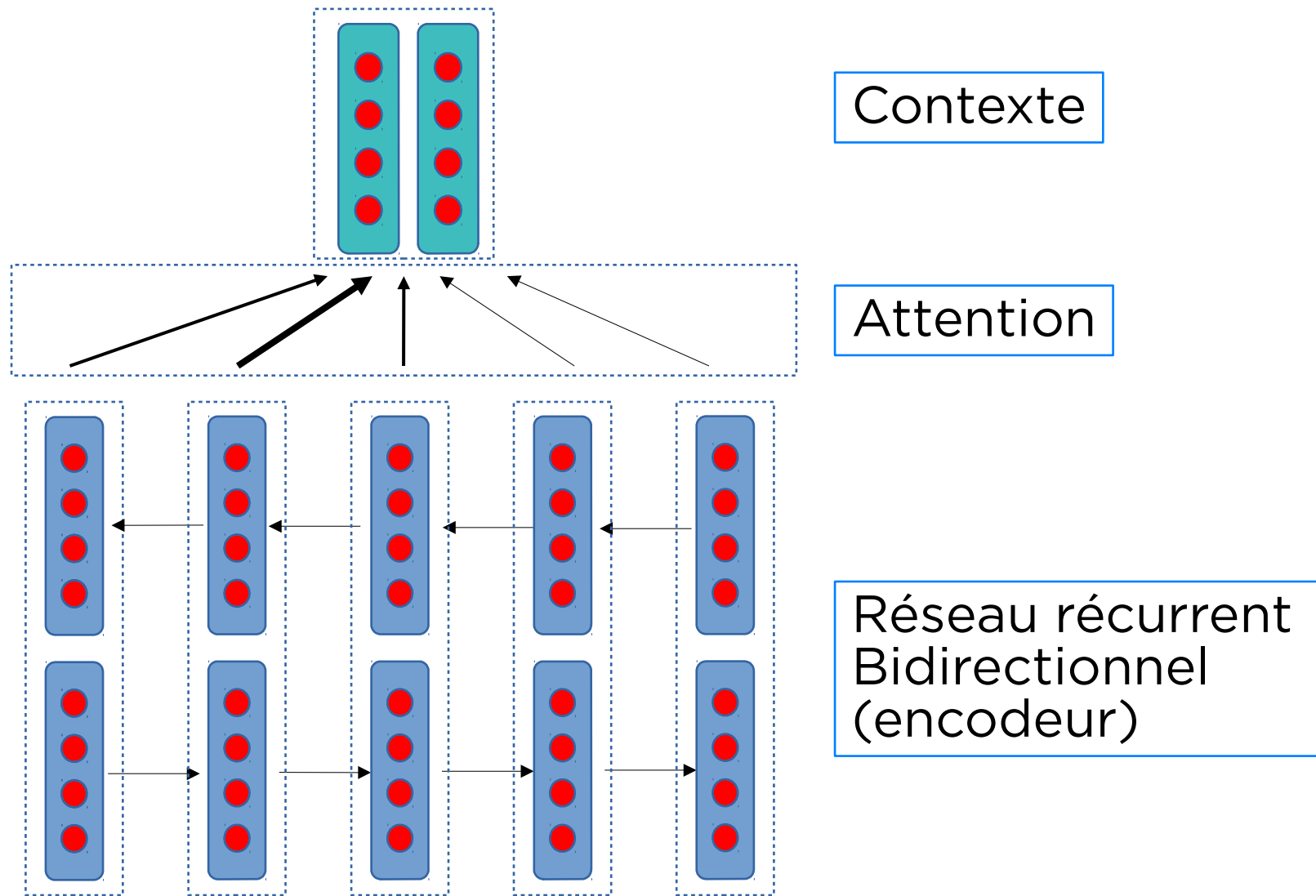
Bahdanau, D., Cho, K., & Bengio, Y. Neural machine translation by jointly learning to align and translate. ICLR (2015)

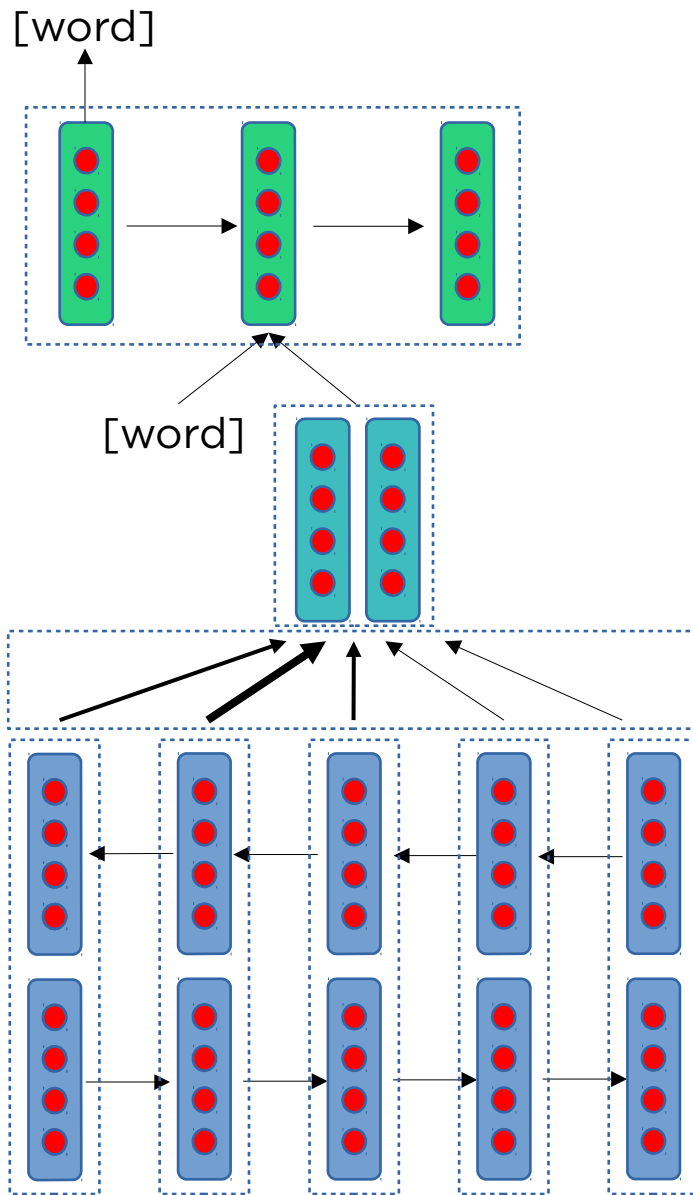


Bahdanau, D., Cho, K., & Bengio, Y.  
 Neural machine translation  
 by jointly learning to align  
 and translate. ICLR (2015)



Réseau récurrent  
Bidirectionnel  
(encodeur)



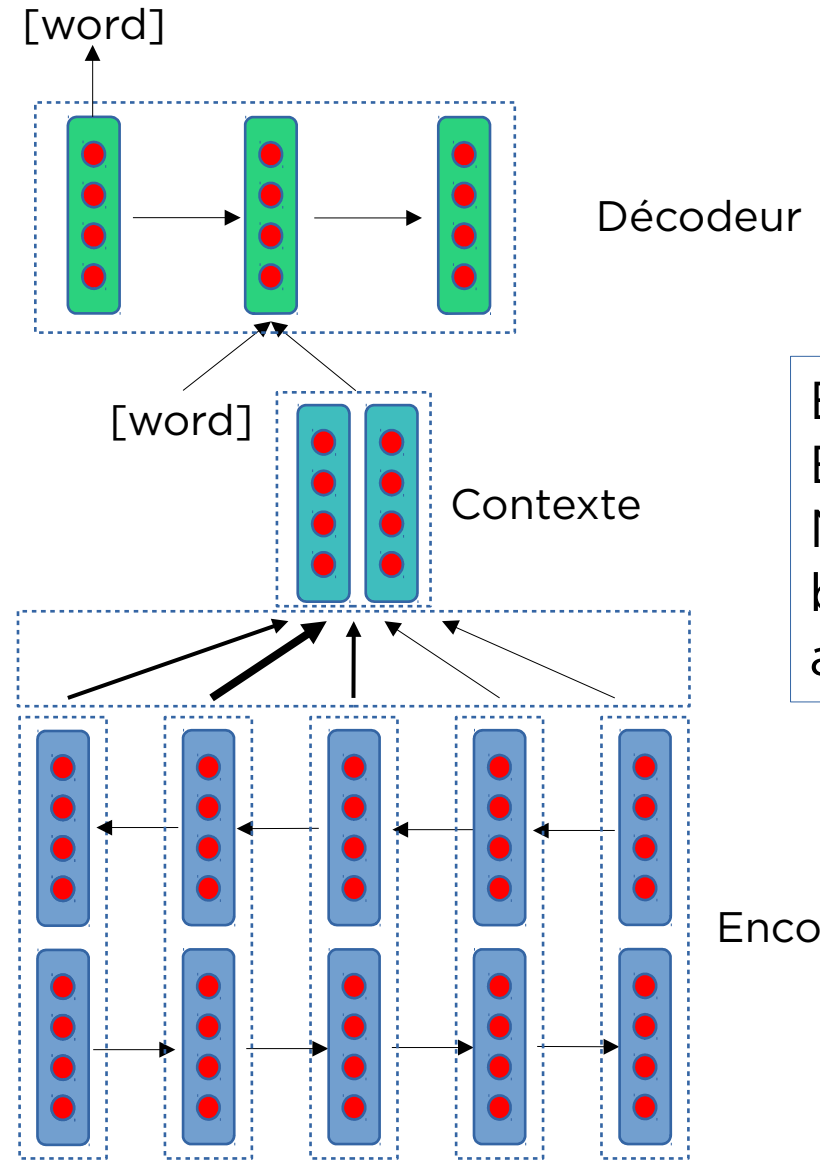
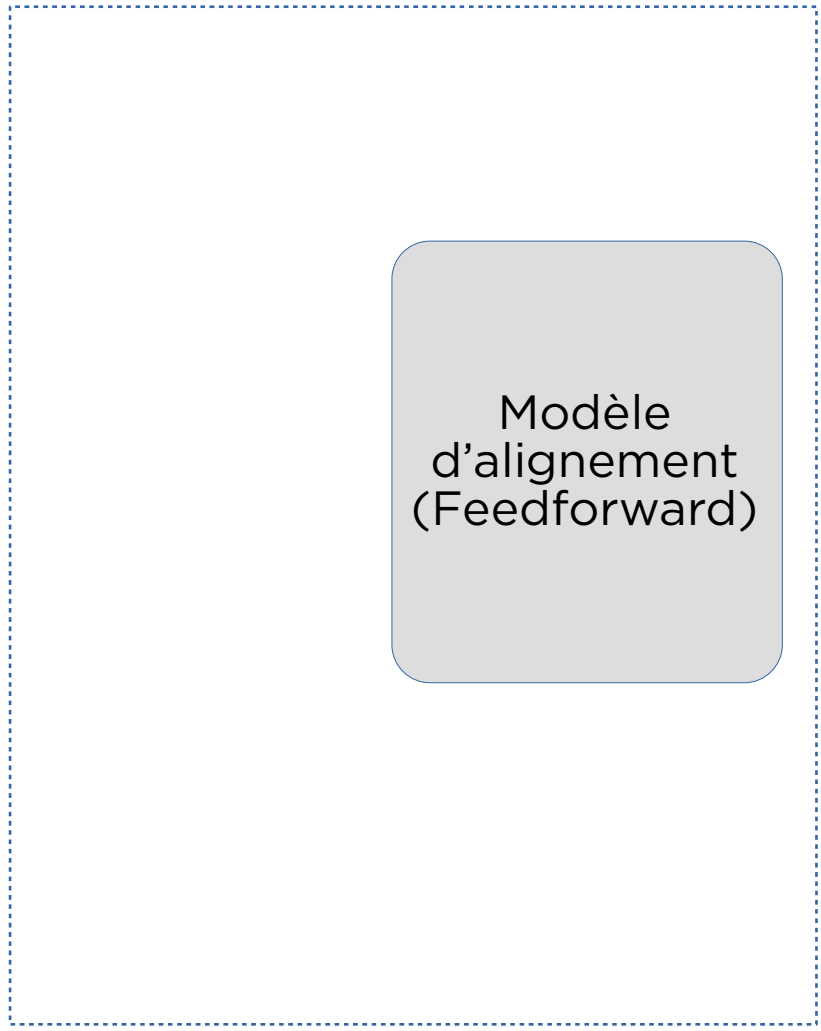


Réseau récurrent  
(décodeur)

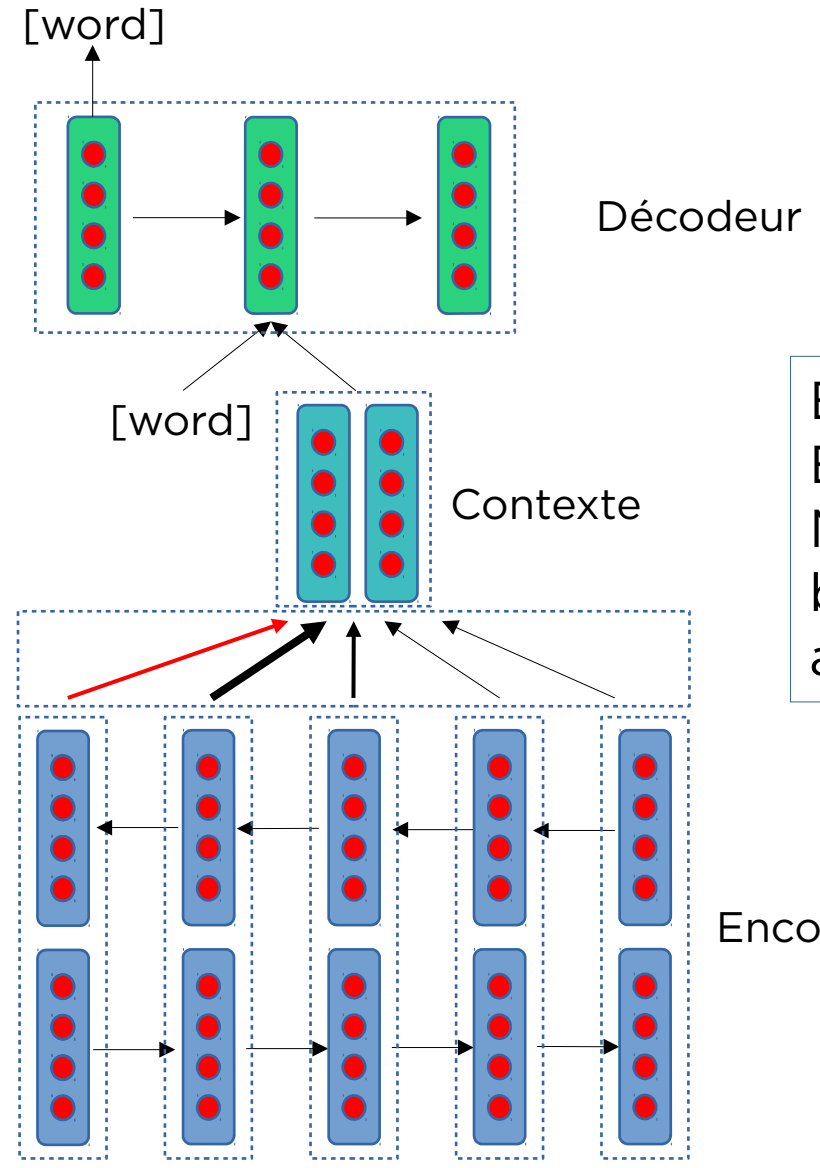
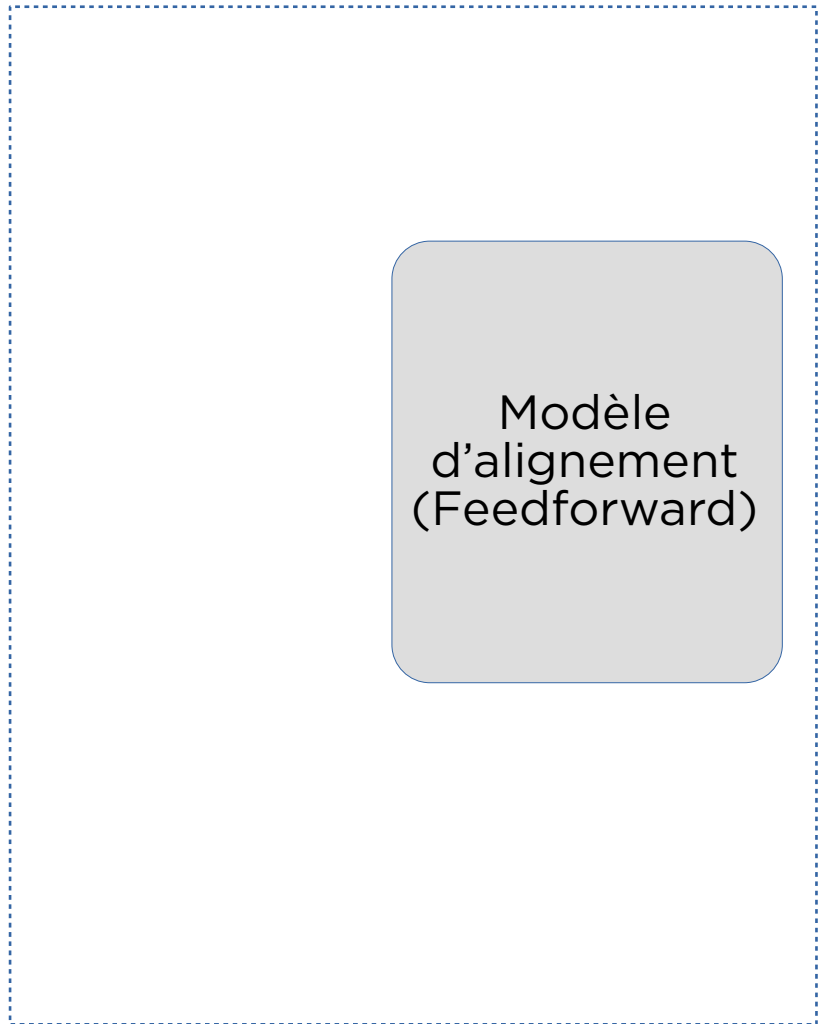
Contexte

Attention

Réseau récurrent  
Bidirectionnel  
(encodeur)

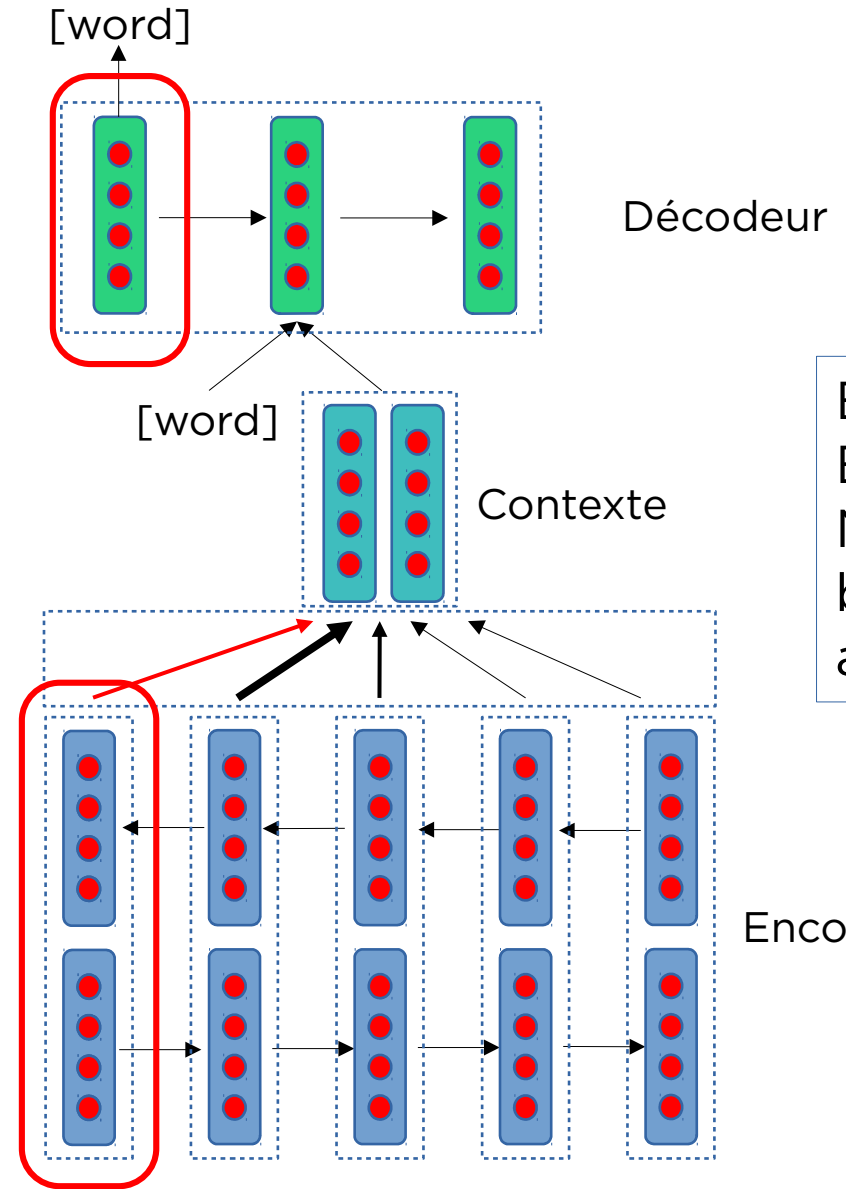
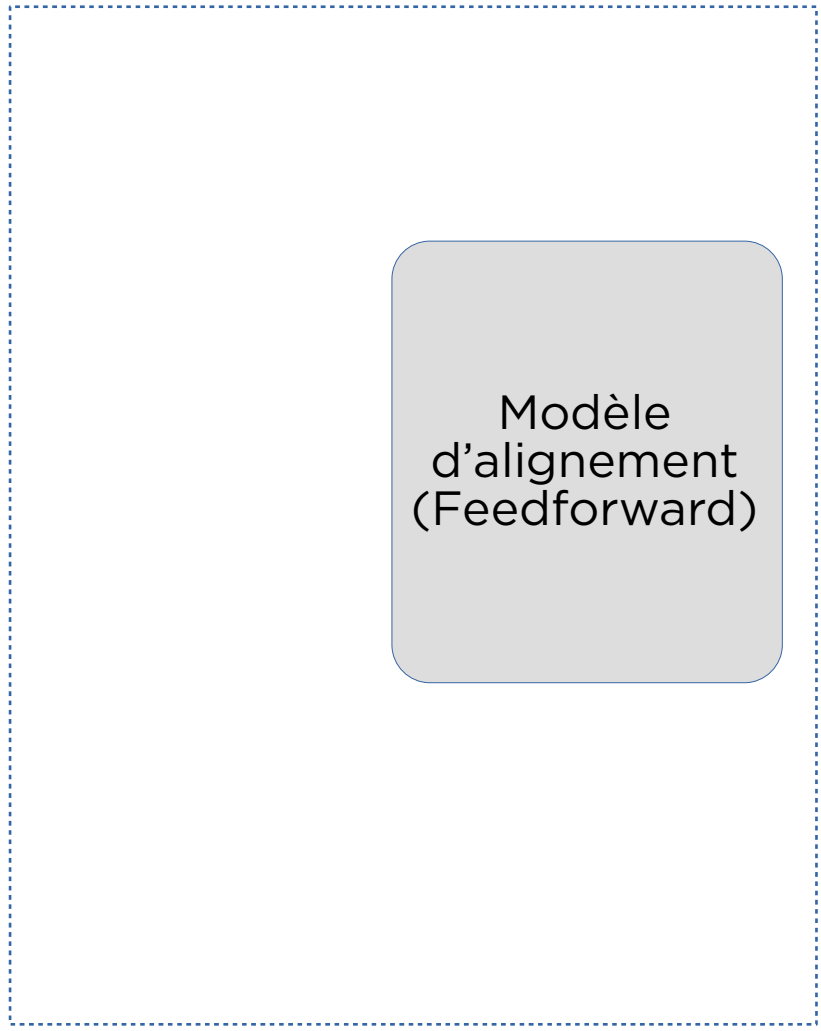


Bahdanau, D., Cho, K., & Bengio, Y.  
Neural machine translation  
by jointly learning to align  
and translate. ICLR (2015)

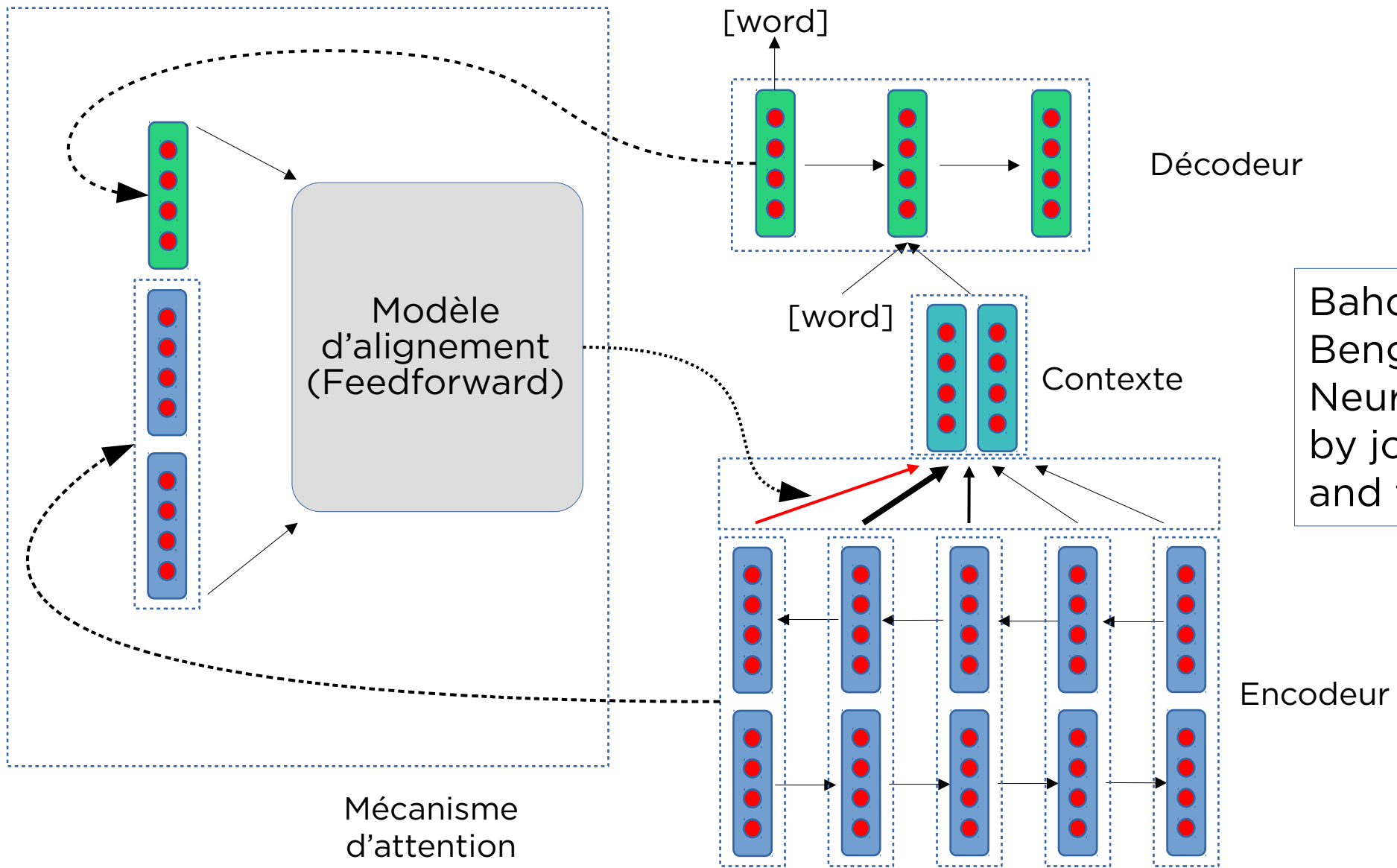


Bahdanau, D., Cho, K., & Bengio, Y.  
Neural machine translation  
by jointly learning to align  
and translate. ICLR (2015)

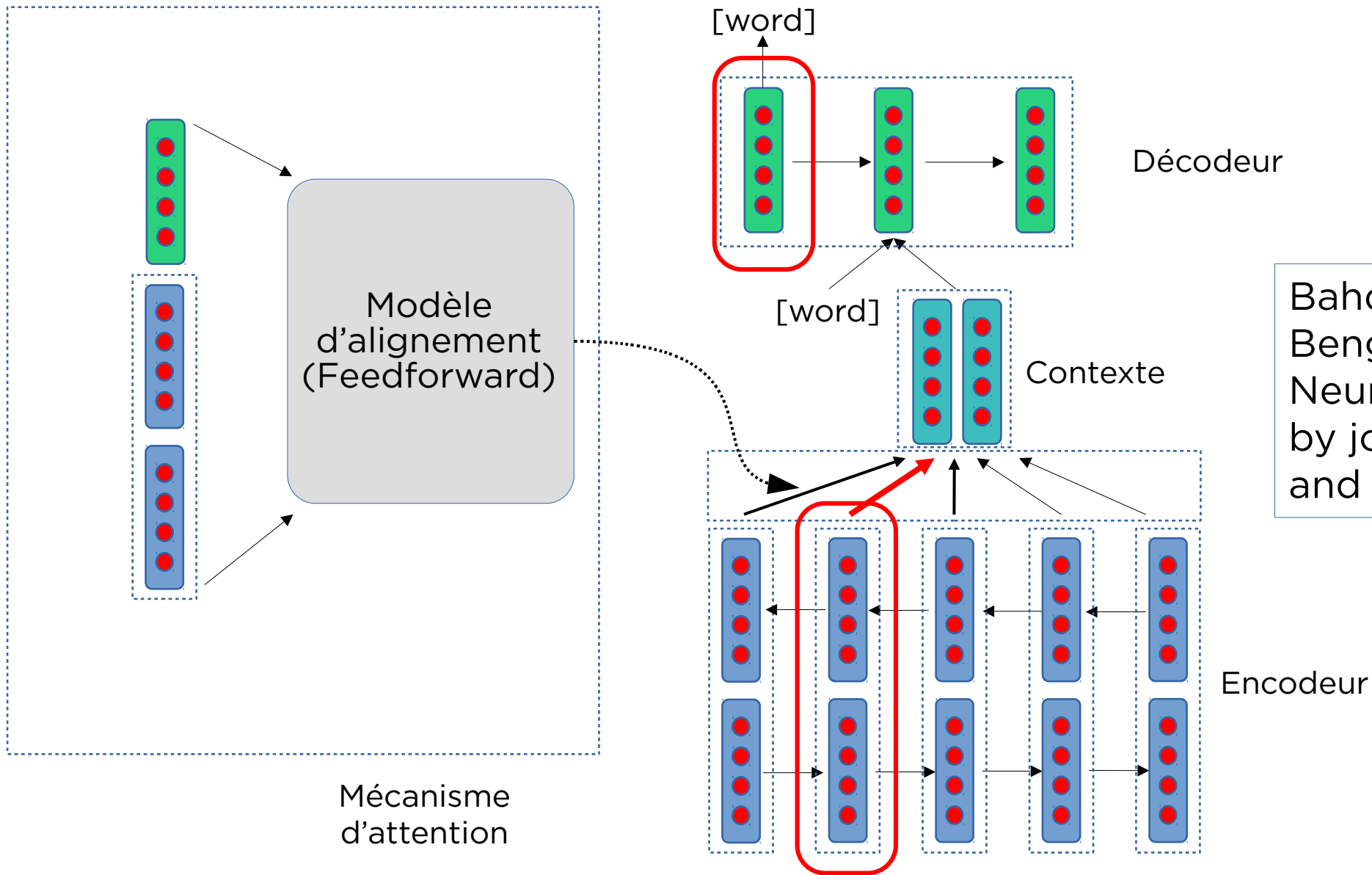




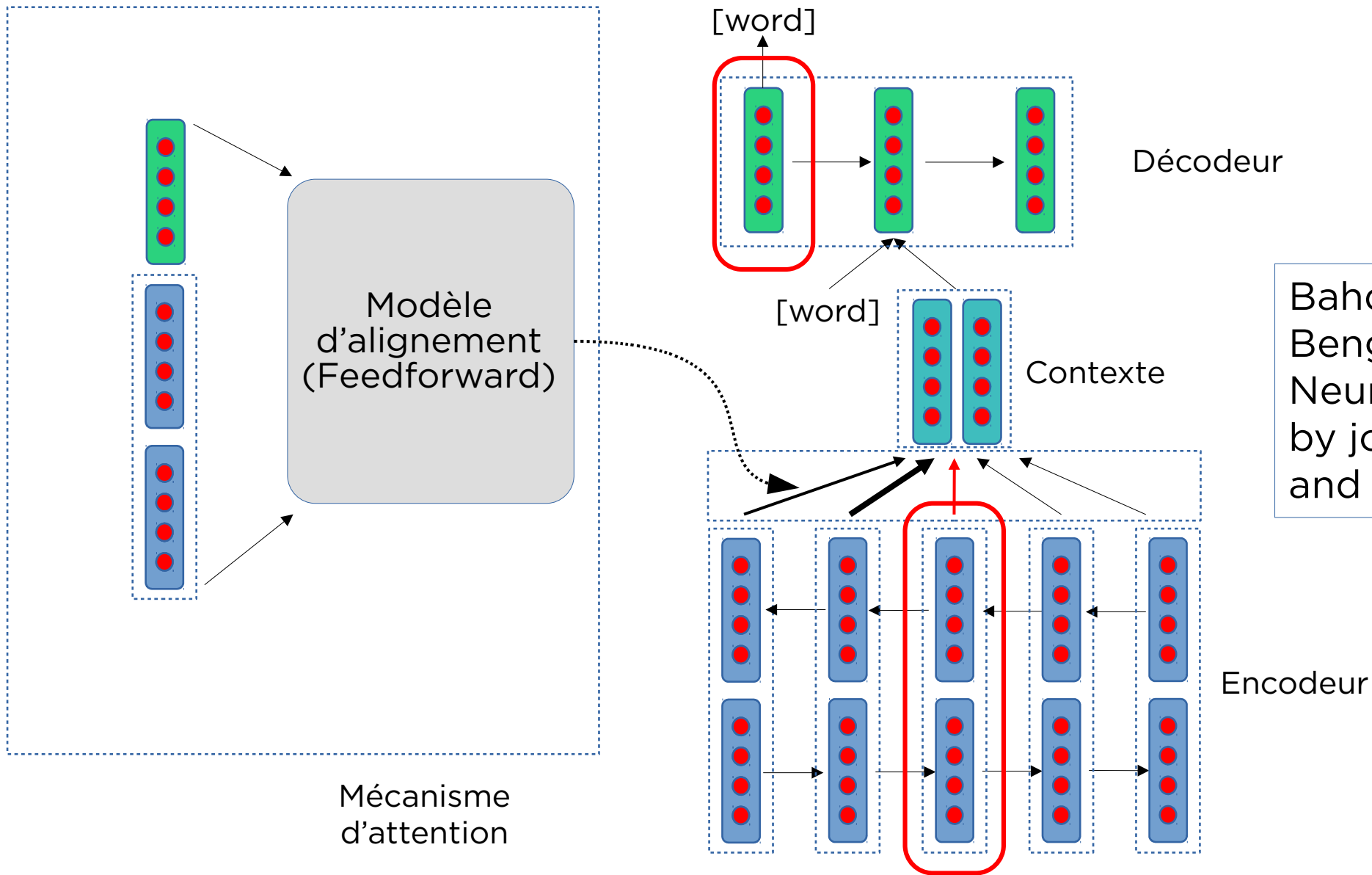
Bahdanau, D., Cho, K., & Bengio, Y.  
Neural machine translation by jointly learning to align and translate. ICLR (2015)



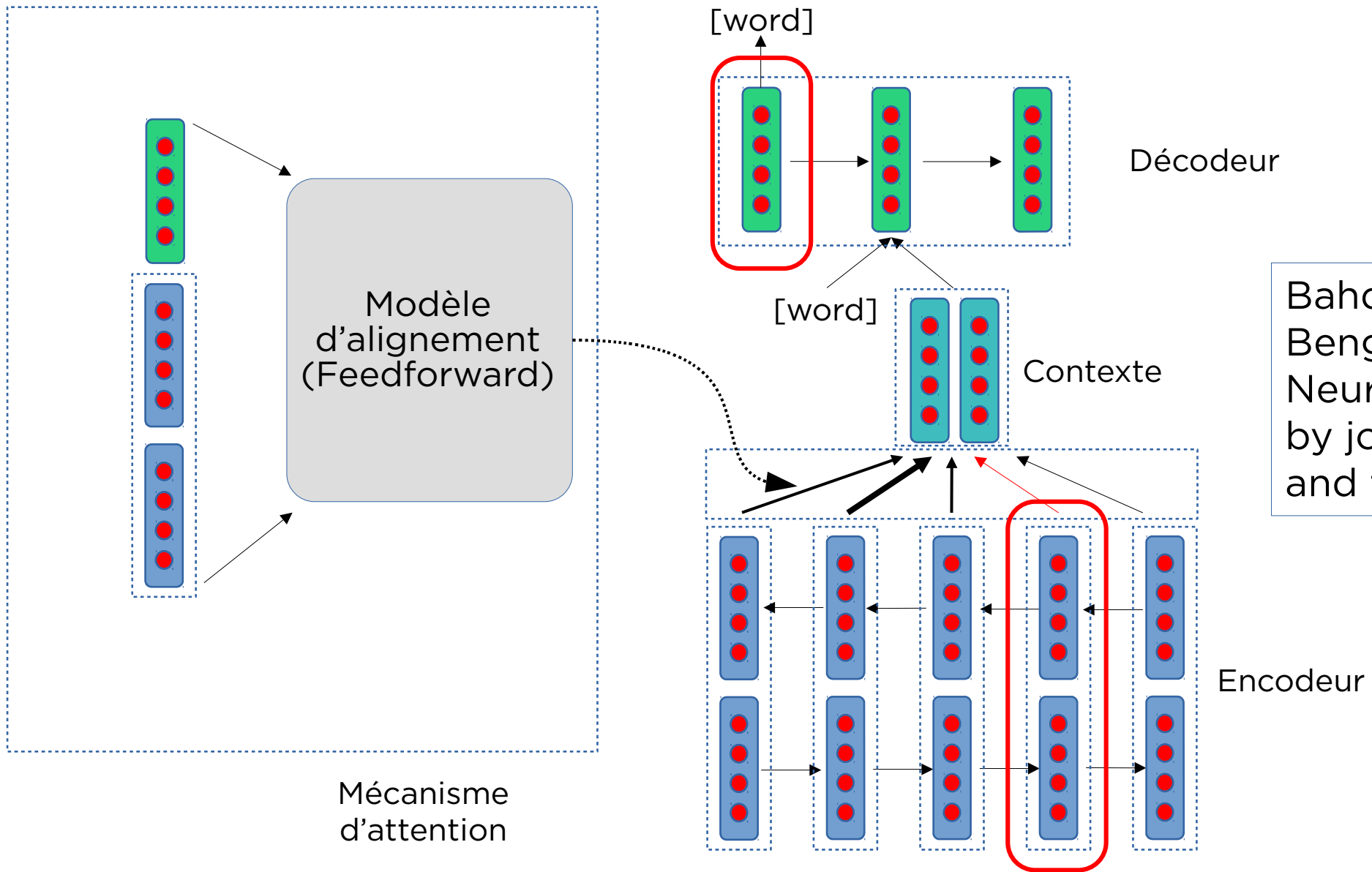
Bahdanau, D., Cho, K., & Bengio, Y.  
 Neural machine translation  
 by jointly learning to align  
 and translate. ICLR (2015)



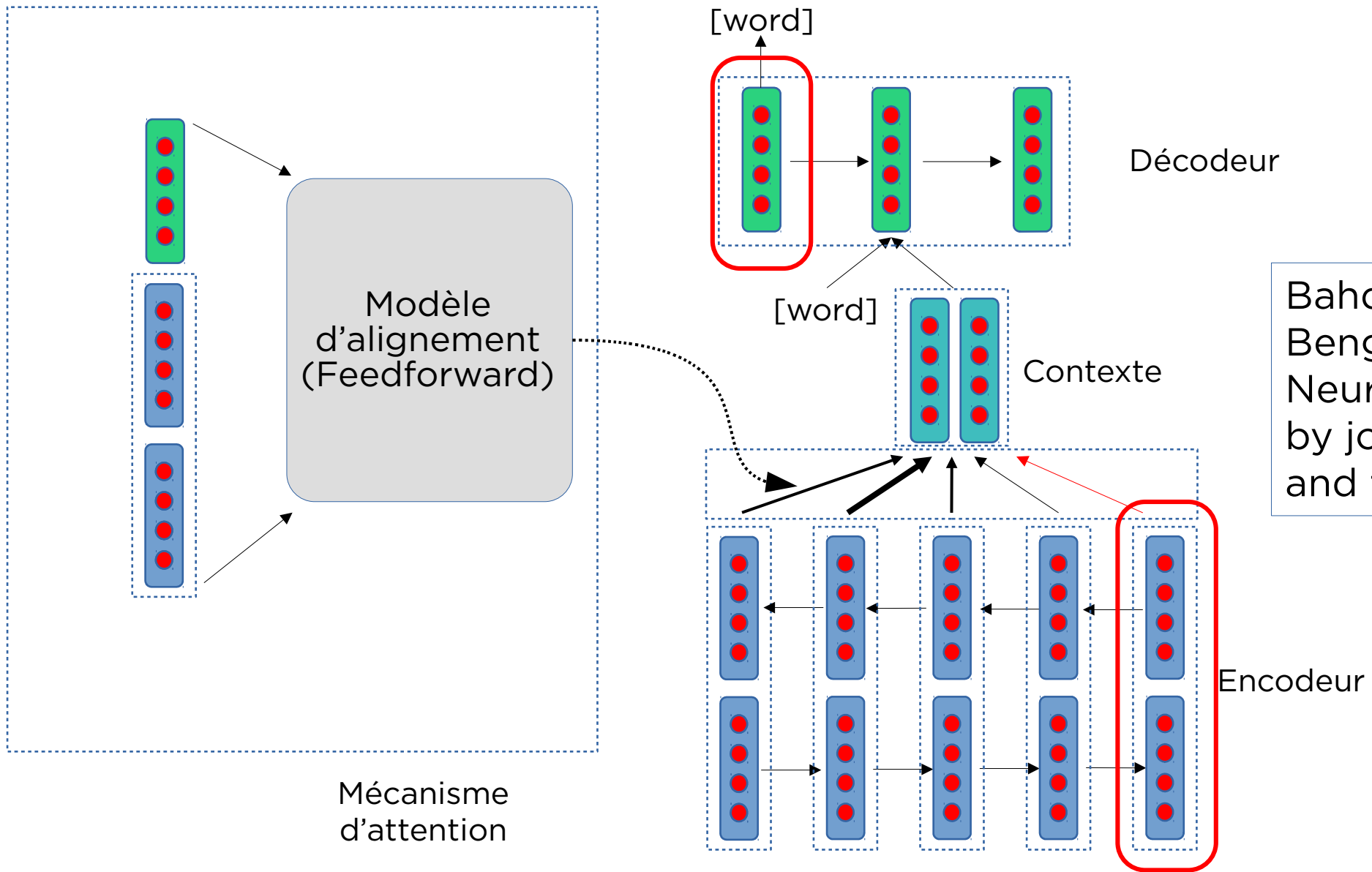
Bahdanau, D., Cho, K., & Bengio, Y.  
 Neural machine translation  
 by jointly learning to align  
 and translate. ICLR (2015)



Bahdanau, D., Cho, K., & Bengio, Y.  
 Neural machine translation  
 by jointly learning to align  
 and translate. ICLR (2015)



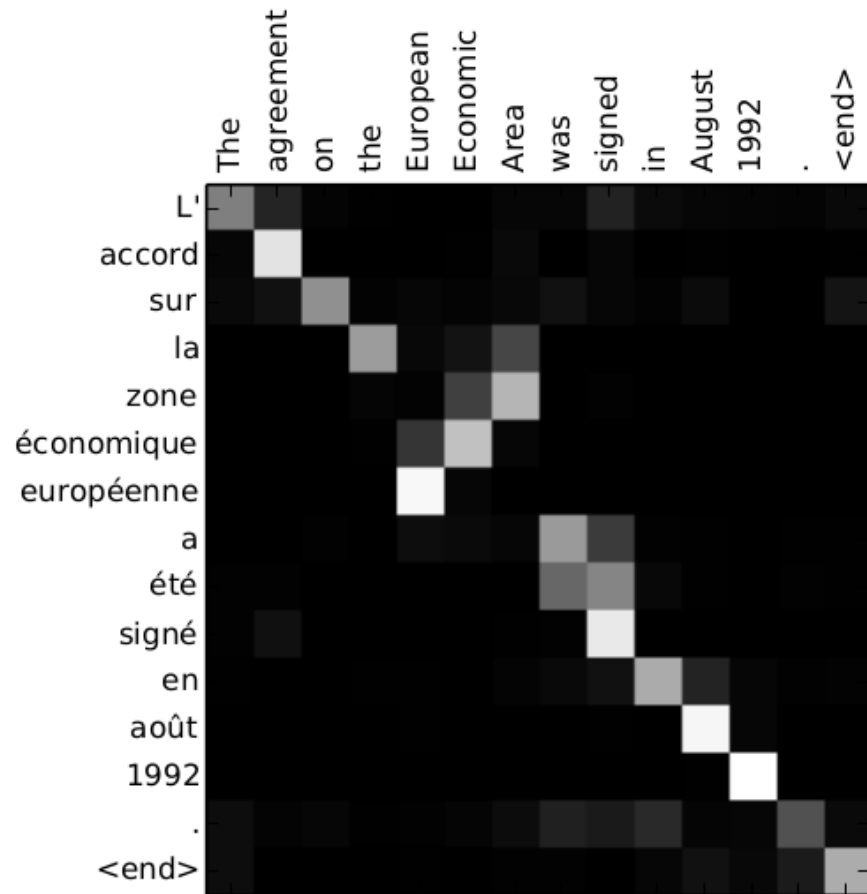
Bahdanau, D., Cho, K., & Bengio, Y.  
 Neural machine translation  
 by jointly learning to align  
 and translate. ICLR (2015)



Bahdanau, D., Cho, K., & Bengio, Y.  
 Neural machine translation  
 by jointly learning to align  
 and translate. ICLR (2015)

# Visualiser l'attention

Bahdanau, D., Cho, K., & Bengio, Y.  
Neural machine translation by  
jointly learning to align  
and translate. ICLR (2015)



(a)

Autre visualisation:

<https://distill.pub/2016/augmented-rnns/#attentional-interfaces>

# En résumé...

- Deux hypothèses de travail validées:
  - Hypothèse distributionnelle → Représentation distribuée
  - Compositionnalité → Opération sur les représentations distribuées



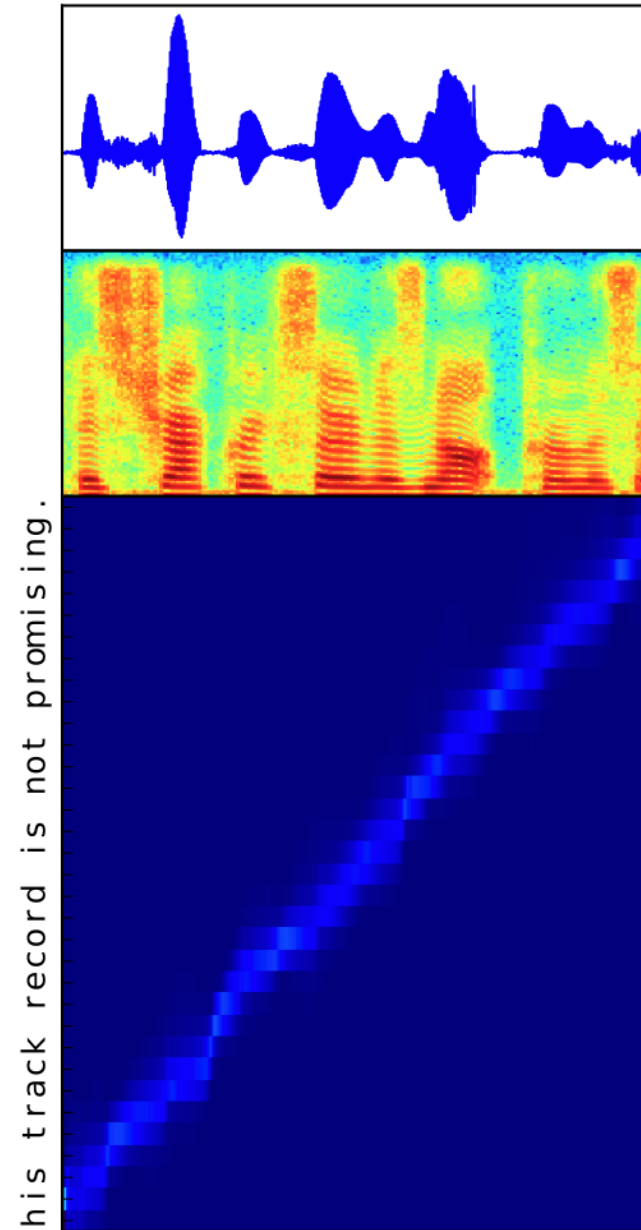
# En résumé...

- Le NLP est un domaine en soi avec beaucoup d'applications et de techniques existantes
- Le Deep Learning est un outil essentiel dans plusieurs tâches du NLP
- Il reste encore de la recherche à faire sur des problèmes complexes, e.g., Q&A et chatbot

# Au-delà du NLP

Convertir des types  
de données:  
Texte ↔ Discours parlé

Sotelo, J., Mehri, S., Kumar, K., Santos, J. F.,  
Kastner, K., Courville, A., & Bengio, Y. "Char2Wav:  
End-to-end speech synthesis." (2017).



# Au-delà du NLP

- GuessWhat ?





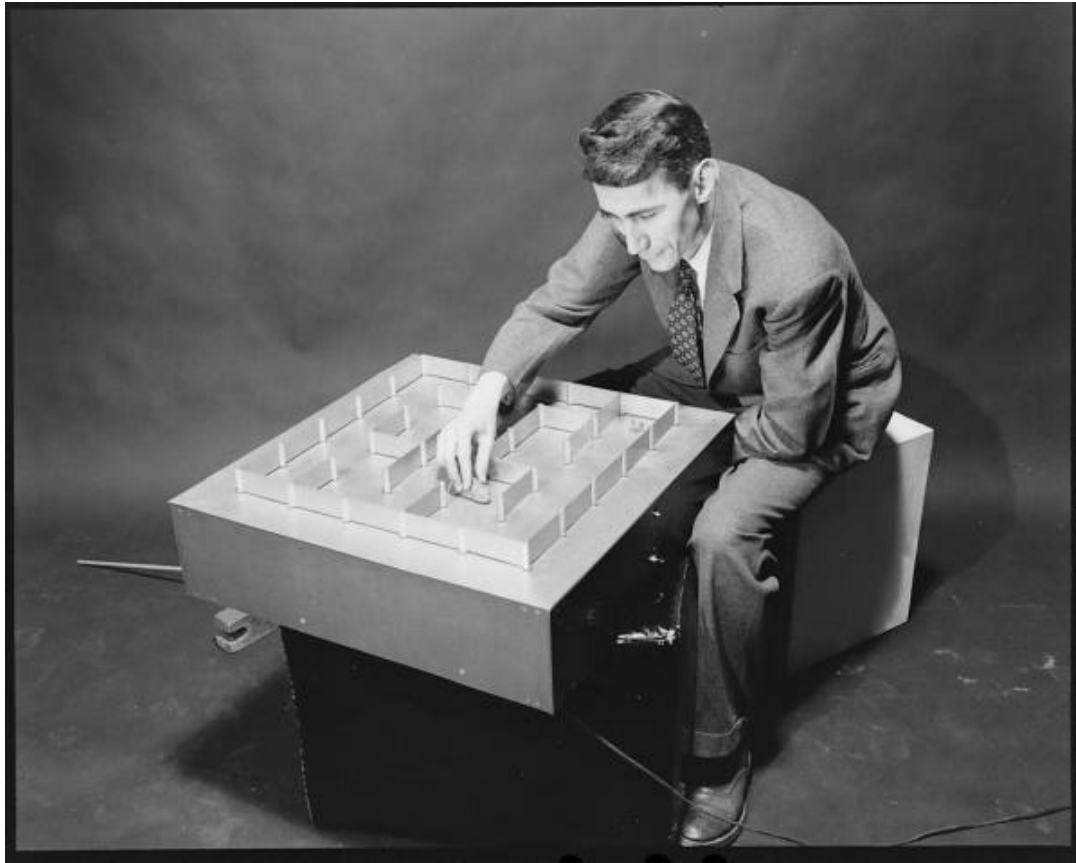
Image	Beam Search	REINFORCE	Image	Beam Search	REINFORCE
	Is it a person ? no Is it a ball ? no Is it a ball ? no Is it a ball ? no Is it a ball ? no Is it a ball ? no <b>Failure</b> (blue bat)	Is it a person ? no Is a glove ? no Is an umbrella ? no Is in the middle ? no On a person? no is it on on far right? yes <b>Success</b> (red chair)		Is it a cat ? no Is it a book ? no Is it a book ? no Is it a book ? no Is it a book ? no Is it a book ? no <b>Failure</b> (person)	Is it a cat ? no Is it on the table ? yes Is it the book ? no Is it fully visible? yes <b>Success</b> (bowl)
	Is it a person ? yes Is it the one in front ? yes Is it the one on the left ? no Is it the one in the middle with the red umbrella ? yes Is it the one to the right of the girl in ? no <b>Failure</b> (umbrella)	Is it a person ? yes Is it girl in white ? yes <b>Success</b> (girl)		Is it a bag ? yes Is it red ? no Is it the one in the middle ? no Is it the one on the far right ? no Is it the one with the blue bag ? yes <b>Success</b> (most left bag)	Is it a suitcase? yes Is it in the left side ? yes <b>Failure</b> (left bag)

Table 1: Samples extracted from the test set. The blue (resp. purple) box corresponds to the object picked by the guesser for the beam-search (resp. REINFORCE) dialogue. The small verbose description is added to refer to the object picked by the guesser.

Strub, Florian, et al. "End-to-end optimization of goal-driven and visually grounded dialogue systems." International Joint Conference on Artificial Intelligence. 2017.

# Qu'y a-t-il dans la machine ?



Institut  
des algorithmes  
d'apprentissage  
de Montréal



**Merci pour votre attention!**

# Références

- A mathematical theory of communication Bell system technical journal, Vol. 27 (1948) by C. E. Shannon
- Radford, A., Jozefowicz, R., & Sutskever, I. (2017). Learning to generate reviews and discovering sentiment. preprint arXiv:1704.01444. (<https://blog.openai.com/unsupervised-sentiment-neuron/>)
- Bahdanau, D., Cho, K., & Bengio, Y. Neural machine translation by jointly learning to align and translate. ICLR (2015)
- IBM Watson: <https://www.forbes.com/sites/jasonbloomberg/2017/07/02/is-ibm-watson-a-joke/#26a5c035da20>
- Harris, Zellig S. "Distributional structure." Papers in structural and transformational linguistics. Springer Netherlands, 1970. 775-794.
- Bengio, Yoshua, et al. "A neural probabilistic language model." Journal of machine learning research 3.Feb (2003): 1137-1155.
- Collins, Michael. "Log-Linear Models", (<http://www.cs.columbia.edu/~mcollins/loglinear.pdf>)
- Mikolov, Tomas, et al. "Distributed representations of words and phrases and their compositionality." Advances in neural information processing systems. 2013.
- Rong, Xin. "word2vec parameter learning explained." arXiv preprint arXiv:1411.2738 (2014).
- Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space." arXiv preprint arXiv:1301.3781 (2013).
- Kyunghyun Cho, Bart Merriënboer, Caglar Gulcehre, Fethi Bougares, Holger Schwenk, et al.. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. EMNLP, 2014, Unknown, Unknown or Invalid Region. 2014.
- Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." Advances in neural information processing systems. (2014).
- Sotelo, J., Mehri, S., Kumar, K., Santos, J. F., Kastner, K., Courville, A., & Bengio, Y. "Char2Wav: End-to-end speech synthesis." (2017).
- Xu, Kelvin, et al. "Show, attend and tell: Neural image caption generation with visual attention." International Conference on Machine Learning. 2015.
- Strub, Florian, et al. "End-to-end optimization of goal-driven and visually grounded dialogue systems." International Joint Conference on Artificial Intelligence. 2017.