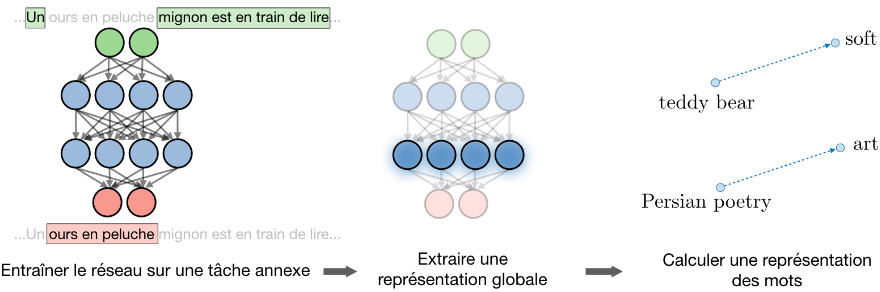
2.3.2 Représentation de mots

❒ **Word2vec** – Word2vec est un ensemble de techniques visant à apprendre comment représenter les mots en estimant la probabilité qu'un mot donné a d'être entouré par d'autres mots. Le skip-gram, l'échantillonnage négatif et le CBOW font parti des modèles les plus populaires.



❒ **Skip-gram** – Le skip-gram est un modèle de type supervisé qui apprend comment représenter les mots en évaluant la probabilité de chaque mot cible t donné dans un mot contexte c. En notant θ_t le paramètre associé à t, la probabilité $P(t|c)$ est donnée par :

$$\boxed{P(t|c)=\frac{\exp(\theta_t^Te_c)}{\displaystyle\sum_{j=1}^{|V|}\exp(\theta_j^Te_c)}}$$

Remarque : le fait d'additionner tout le vocabulaire dans le dénominateur du softmax rend le modèle coûteux en temps de calcul. CBOW est un autre modèle utilisant les mots avoisinants pour prédire un mot donné.

❒ **Échantillonnage négatif** – Cette méthode utilise un ensemble de classifieurs binaires utilisant des régressions logistiques qui visent à évaluer dans quelle mesure des mots contexte et cible sont susceptible d'apparaître simultanément, avec des modèles étant entraînés sur des ensembles de k exemples négatifs et 1 exemple positif. Étant donnés un mot contexte c et un mot cible t, la prédiction est donnée par :

$$\boxed{P(y=1|c,t)=\sigma(\theta_t^Te_c)}$$

Remarque : cette méthode est moins coûteuse en calcul par rapport au modèle skip-gram.

❒ **GloVe** – Le modèle GloVe (en anglais *global vectors for word representation*) est une technique de représentation des mots qui utilise une matrice de co-occurrence X où chaque $X_{i,j}$ correspond au nombre de fois qu'une cible i se produit avec un contexte j. Sa fonction de coût J est telle que :

$$\boxed{J(\theta)=\frac{1}{2}\sum_{i,j=1}^{|V|}f(X_{ij})(\theta_i^Te_j+b_i+b_j'-\log(X_{ij}))^2}$$

où f est une fonction à coefficients telle que $X_{i,j}=0\Longrightarrow f(X_{i,j})=0$.

Étant donné la symmétrie que e et θ ont dans un modèle, la représentation du mot final $e_w^{(\textrm{final})}$ est donnée par :

$$\boxed{e_w^{(\textrm{final})}=\frac{e_w+\theta_w}{2}}$$