

Motivation

- Un nombre croissant d'intérêts de recherche sur les langues autochtones, les réalités et les défis au sein de la communauté internationale du TAL
- Défis :
 - De multiples caractéristiques telles que la polysynthèse, la complexité morphologique
 - Variation dialectale avec riche morpho-phonémie
 - Orthographe avec des données éparses et des scénarios de ressources limitées
- **Objectifs de recherche :**
 - Examiner la littérature relative aux systèmes de traduction automatique pour les langues autochtones
 - Construire notre première traduction automatique neuronale inuktitut-anglais en réalisant des fonctionnalités supplémentaires extraites des informations d'alignement source-cible et d'autres lexiques bilingues

Méthodologie

Le but de ce projet de recherche vise à étudier notre premier système de traduction automatique neuronale inuktitut-anglais basé sur l'architecture encodeur-décodeur de Transformer [1]. Notre approche se compose de trois parties principales.

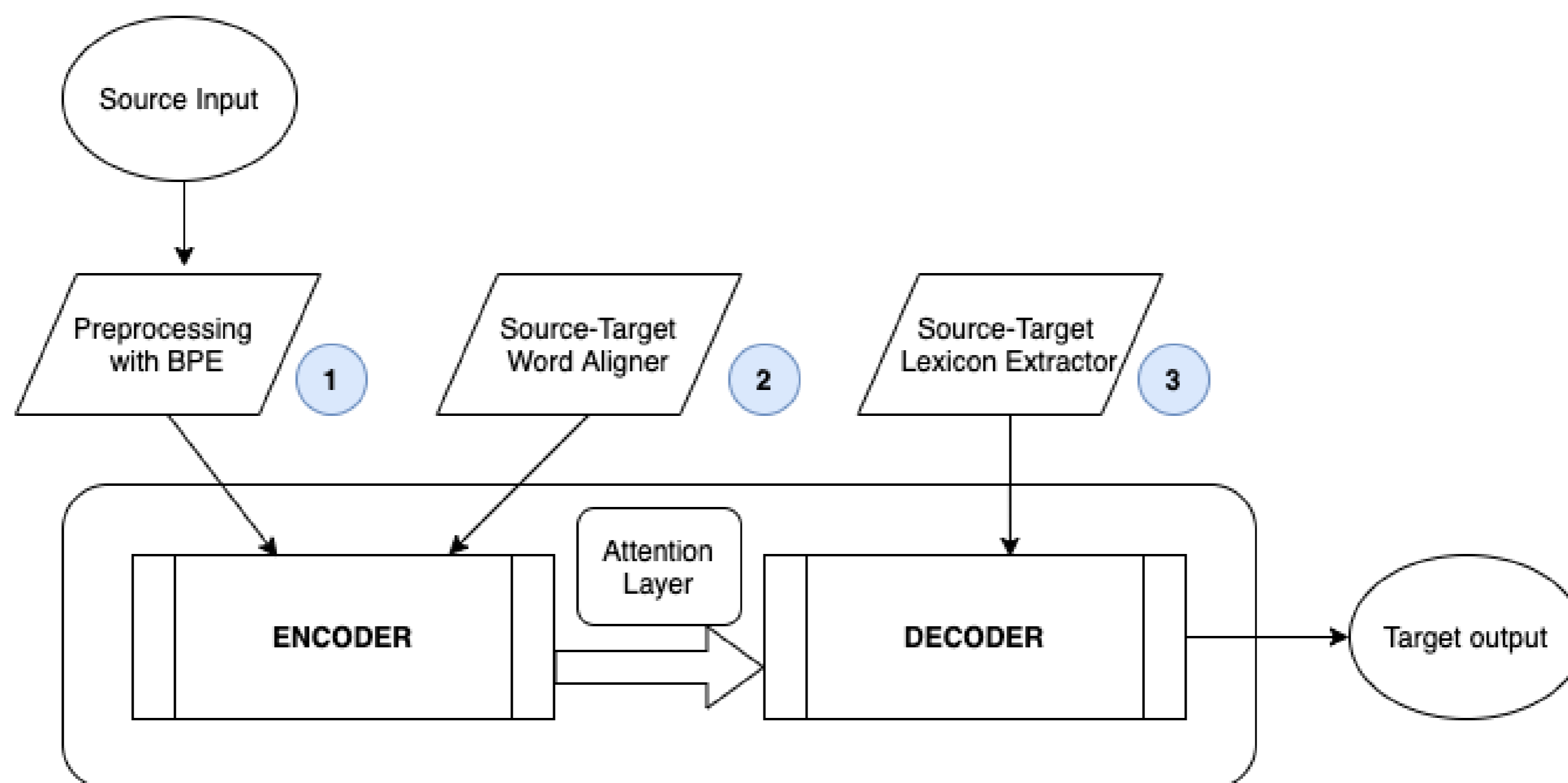


Figure 1. Architecture de notre premier système de traduction automatique neuronale inuktitut-anglais basé sur l'architecture encodeur-décodeur de Transformer

Expérimentations et Évaluation

Inuktitut → Anglais	Données de Test	Inuktitut → Anglais	Données de Test
SMT [2]	27.80	Baseline + align information (Système 1)	35.71
Baseline NMT [2]	35.00	Baseline + lex.s2t (Système 2)	35.93
		All (Système 3)	36.03

Table 1. Comparaison de notre approche proposée pour la TAN basée sur l'architecture de Transformer, en termes de scores BLEU

Conclusion et Perspectives

- Revue des recherches en TAL : Traduction automatique, l'une des technologies linguistiques les plus connues pour revitaliser et préserver les langues autochtones et en danger
- Perspectives :
 - Davantage de recherches pour améliorer le développement de la technologie des langues autochtones
 - Établir une étroite collaboration avec les organisations et les communautés autochtones partout au Canada

Bibliographie

- [1] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.
- [2] Eric Joanis, Rebecca Knowles, Roland Kuhn, Samuel Larkin, Patrick Littell, Chi-kiu Lo, Darlene Stewart, and Jeffrey Micher. 2020. The Nunavut Hansard Inuktitut-English parallel corpus 3.0 with preliminary machine translation results. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 2562–2572, Marseille, France, May. European Language Resources Association.