HAL open science

Revitalisation des langues autochtones via le prétraitement et la traduction automatique neuronale: le cas de l'inuktitut

Tan Le Ngoc, Fatiha Sadat

To cite this version:

Tan Le Ngoc, Fatiha Sadat. Revitalisation des langues autochtones via le prétraitement et la traduction automatique neuronale: le cas de l'inuktitut. Traitement Automatique des Langues Naturelles, 2021, Lille, France. pp.273-274. hal-03265883

HAL Id: hal-03265883

https://hal.science/hal-03265883v1

Submitted on 23 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Actes de la 28e Conférence sur le Traitement Automatique des Langues Naturelles
Lille, France, 28 juin au 2 juillet 2021

Revitalisation et Préservation des Langues Autochtones via le Traitement Automatique de la Langue Naturelle

Ngoc Tan Le Fatiha Sadat
Université du Québec à Montréal / Montréal, Québec, Canada
le.ngoc_tan@uqam.ca, sadat.fatiha@uqam.ca

RÉSUMÉ

Nous présentons des résumés en français et en anglais de l'article (Tan Le & Sadat, 2020) présenté à la 28ème conférence internationale sur les linguistiques computationnelles (*the 28th International Conference on Computational Linguistics*) en 2020.

ABSTRACT

Revitalization and Preservation of Indigenous Languages through Natural Language Processing

We present French and English abstracts of the article (Tan Le & Sadat, 2020) that was presented at the 28th International Conference on Computational Linguistics en 2020.

MOTS-CLÉS : Langues autochtones, inuktitut, langues peu dotées, prétraitement, segmentation morphologique, traduction automatique neuronale.

KEYWORDS: Indigenous languages, Inuktitut, low-resource languages, Preprocessing, Morphological Segmentation, Neural Machine Translation.

1 Résumé en français

Le traitement des langues autochtones a été très difficile au sein des tâches et des applications du TALN pour multiples raisons. En général, ces langues, dans la dimension linguistique, sont polysynthétiques et fortement infléchies avec une morphophonémie riche et des orthographes dialectales variables. De plus, les langues autochtones ont été considérées comme peu dotées et/ou en danger ; ce qui pose un grand défi pour la recherche liée à l'intelligence artificielle et à ses axes de recherche, dont le TALN.

Dans cet article [1], nous proposons une étude sur l'inuktitut afin de revitaliser et préserver la langue qui appartient à la famille inuite, parlée dans le nord du Canada. Nous nous concentrons sur : (1) le prétraitement, et (2) les applications sur des tâches du TALN spécifiques telles que l'analyse morphologique et la traduction automatique neuronale. Nos évaluations dans le contexte de la traduction automatique neuronale inuktitut-anglais peu dotée ont montré des améliorations significatives par rapport à l'état de l'art.

1. Présenté à Coling 2020 : https://www.aclweb.org/anthology/2020.coling-main.410/

Cette œuvre est mise à disposition sous licence Attribution 4.0 International.

2 English Abstract

Indigenous languages have been very challenging when dealing with NLP tasks and applications because of multiple reasons. These languages, in linguistic typology, are polysynthetic and highly inflected with rich morphophonemics and variable dialectal-dependent spellings ; which affected studies on any NLP task in the recent years. Moreover, Indigenous languages have been considered as low-resource and/or endangered ; which poses a great challenge for research related to Artificial Intelligence and its fields, such as NLP.

In this paper [2], we propose a study on the Inuktitut through pre-processing and neural machine translation, in order to revitalize the language which belongs to the Inuit family, spoken in Northern Canada. Our focus is concentrated on : (1) the preprocessing, and (2) applications on specific NLP tasks such as morphological analysis and neural machine translation. Our evaluations in the context of low-resource Inuktitut-English Neural Machine Translation, showed significant improvements of the proposed approach compared to the state-of-the-art.

Remerciements

The authors wish to express their thanks for professor Richard Compton, Canada Research Chair in Knowledge and Transmission of the Inuit Language, and the reviewers, for their constructive feedbacks.

Références

TAN LE N. & SADAT F. (2020). Revitalization of indigenous languages through pre-processing and neural machine translation : The case of Inuktitut. In *Proceedings of the 28th International Conference on Computational Linguistics*, p. 4661–4666, Barcelona, Spain (Online) : International Committee on Computational Linguistics. DOI : 10.18653/v1/2020.coling-main.410.

2. Presented at Coling 2020 : https://www.aclweb.org/anthology/2020.coling-main.410/