Traduire des corpus pour construire des modèles de traduction neuronaux : une solution pour toutes les langues peu dotées ?

Raoul Blin

Actes de la 6e conférence conjointe Journées d'Études sur la Parole (JEP, 31e édition),
Traitement Automatique des Langues Naturelles (TALN, 27e édition),
Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition)
Nancy, France, 08-19 juin 2020

Traduire des corpus pour construire des modèles de traduction neuronaux : une solution pour toutes les langues peu dotées ?

Raoul Blin
CNRS-CRLAO, 105 Bd Raspail, 75006 Paris, France
blin@ehess.fr

RÉSUMÉ

Nous comparons deux usages des langues pivots en traduction automatique neuronale pour des langues peu dotées. Nous nous intéressons au cas où il existe une langue pivot telle que les paires source-pivot et pivot-cible sont bien ou très bien dotées. Nous comparons la traduction séquentielle traditionnelle (source→pivot→cible) et la traduction à l'aide d'un modèle entraîné sur des corpus traduits à l'aide des langues pivot et cible. Les expériences sont menées sur trois langues sources (espagnol, allemand et japonais), une langue pivot (anglais) et une langue cible (français). Nous constatons que quelle que soit la proximité linguistique entre les langues source et pivot, le modèle entraîné sur corpus traduit a de meilleurs résultats que la traduction séquentielle, et bien sûr que la traduction directe.

ABSTRACT

Corpus Translation to Build Translation Models : a Solution for all Low-Resource Languages ?

We compare two uses of pivot languages in neural machine translation when a language pair is low or not resourced, but there is a pivot language such that the source-pivot and pivot-target pairs are well or very well resourced. We compare traditional sequential translation (source→pivot→target) and translation using a model trained on corpora translated with the pivot and target languages. Experiments conducted with three source languages (Spanish, German and Japanese), a pivot language (English) and a target language (French), show that the trained model on translated corpus has better results than sequential translation, and of course than direct translation.

MOTS-CLÉS : langues peu dotées, traduction automatique neuronale, langue pivot.

KEYWORDS: low-resource languages, neural machine translation, pivot language .

1 Introduction

Pour profiter des avantages de la traduction automatique neuronale, il est nécessaire de disposer de très volumineux corpus (Koehn & Knowles, 2017). Malheureusement, peu de paires de langues disposent de telles ressources. Une première solution très étudiée actuellement est le recours à des modèles multilingues de traduction (Rikters *et al.*, 2018), etc.). L'inconvénient est que l'entraînement nécessite plus de données (grand corpus) et plus de puissance de calcul. Cette solution s'inscrit d'ailleurs dans une tendance générale observable tant dans le domaine académique que dans l'industrie, d'un intérêt croissant pour des systèmes toujours plus puissants (Aharoni *et al.*, 2019; Arivazhagan *et al.*, 2019). Pourtant, il existe aussi un besoin pour des systèmes «frugaux» avec de faibles capacités de calcul. Il est alors important de limiter la taille des corpus.

Volume 2 : Traitement Automatique des Langues Naturelles, pages 172–180.
HAL : hal-02784765.

Il existe quelques alternatives à cette course à la puissance pour traiter les paires de langues peu dotées. Une première piste récemment proposée est l'apprentissage par transfert de modèles existants (Kocmi & Bojar, 2019). Une autre solution est le recours aux langues pivots, technique ancienne en traduction mais qui revient sous de nouvelles formes grâce à la traduction automatique neuronale. La pratique traditionnelle consiste à traduire séquentiellement, de la langue source vers la langue pivot puis vers la langue cible. C'est une méthode encore appliquée même dans les systèmes commerciaux (Blin, 2018b). L'inconvénient majeur de la traduction séquentielle automatique avec des systèmes qui restent encore imparfaits, c'est que des informations peuvent être perdues lors de la traduction de la source vers le pivot.

Pour résoudre ce problème, (Currey & Heafield, 2019) ont proposé d'utiliser le pivot pour traduire les corpus eux-mêmes et produire des corpus synthétiques sur lesquels des modèles de traduction directe sont entraînés. Avant de décrire la procédure, nous convenons des abréviations suivantes. a, a' et a^i désignent des corpus monolingues pour une langue donnée «a». a^n et a^m désignent des corpus distincts d'une même langue. $m(a, b)$ désigne un modèle entraîné sur un corpus aligné a-b pour traduire de a vers b. «src», «piv» et «tgt» désignent respectivement les langues source, pivot et cible. La procédure est la suivante. Elle n'est envisageable que si il existe deux corpus «de base» (non traduits) alignés piv^1-src et piv^2-tgt, et un corpus monolingue piv^0. Supposons que l'on veuille traduire d'une langue source src vers une langue cible tgt. On génère un corpus src' en langue source en traduisant piv^0 grâce au modèle $m(piv^1, src)$ (voir Fig.1). En parallèle, on génère un corpus tgt' en langue cible en traduisant ce même corpus piv^0 à l'aide du modèle $m(piv^2, tgt)$. On dispose ainsi d'un corpus synthétique aligné bilingue src'-tgt'. Il est utilisé pour entraîner un modèle de traduction directe de la source vers la cible (par abus de langage et pour faire simple, nous parlerons de «modèle traduit»). Les auteurs ajoutent au corpus synthétique les corpus originaux piv^1-src et piv^2-tgt et obtiennent ainsi un corpus synthétique et multilingue. Cette technique a été utilisée pour traduire la paire russe-allemand avec l'anglais comme pivot et produit des résultats prometteurs (BLEU ≈ 23). Le premier intérêt est que cette technique fonctionne pour une paire de langue sans corpus aligné. Il devrait donc a fortiori fonctionner pour une paire de langues peu dotée. L'autre intérêt est que le corpus synthétique peut être aussi grand que le corpus monolingue.

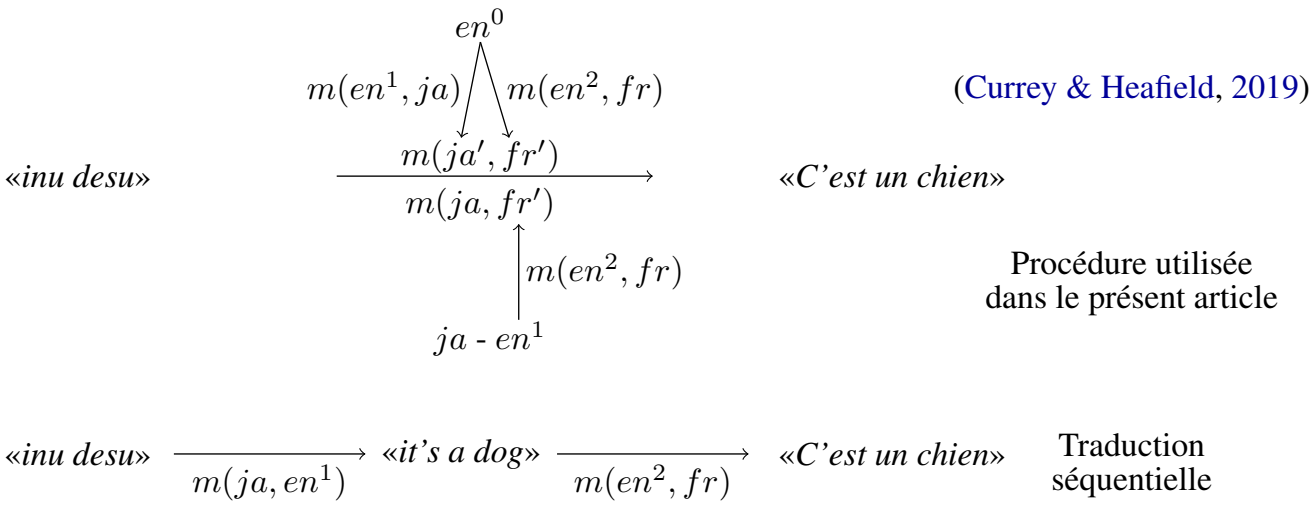

FIGURE 1: Trois procédés de traduction avec langue pivot ; ici, les langues source, pivot et cible sont repectivement le japonais, l'anglais et le français ; $a \xrightarrow{m(x,y)} b$ désigne une traduction de a vers b à l'aide du modèle $m(x, y)$.

Nous nous sommes demandé si cette technique produisait une meilleure traduction que la méthode séquentielle, et dans quelle mesure la proximité entre les langues manipulées influençait les résultats. Pour répondre à cette question, nous avons effectué des comparaisons en jouant sur plusieurs variables : la proximité linguistique des langues source et pivot (les langues pivot et cible restant toujours les mêmes), la part de vocabulaire commun entre les corpus piv^1 et piv^2, et la qualité des traductions. Contrairement à (Currey & Heafield, 2019), nous n'avons pas intégré le multilinguisme car il pouvait affecter différemment la traduction séquentielle et la traduction par modèle traduit.

L'article est organisé en deux parties. Dans un premier temps, nous décrivons les ressources utilisées et la méthode. Puis nous présentons les résultats et les discutons.

2 Expériences

2.1 Langues observées et corpus

Pour prendre en compte la proximité linguistique et la part de vocabulaire commun tout en limitant le nombre d'expériences à mener, nous construisons un ensemble de corpus pour chaque cas extrême : un premier ensemble pour des langues sources proches de la langue pivot et un maximum de vocabulaire en commun, et un second ensemble avec une langue source éloignée et un moindre volume de vocabulaire en commun.

Nous menons les observations pour trois langues sources : espagnol (es), allemand (al) et japonais (ja). La langue cible est le français. La langue pivot est l'anglais. L'anglais a été choisi pour des questions pratiques : il existe un corpus moyen ou grand pour toutes les paires de langues incluant l'anglais et les langues ci-dessus. Les langues sources sont choisies de sorte que la proximité linguistique avec la langue pivot soit progressive. La plus éloignée est le japonais (de type SOV avec marqueurs casuels, pro-drop, non indo-européen, langue agglutinante, écriture différente). L'espagnol est la plus proche (langue SVO, indo-européenne, non agglutinante). L'allemand se situe entre les deux (partiellement SOV, indo-européenne, non agglutinante).

Nous constituons tout d'abord un corpus pour entraîner un modèle de traduction directe, de la source vers la cible. La paire de langues ja-fr est réellement peu dotée ($\approx$ 50K phrases.) (Blin, 2018a). {al,es}-fr sont moyennement ou même bien dotées. Pour simuler des paires faiblement dotées avec l'espagnol et l'allemand, nous utilisons des corpus d'approximativement la même taille que le corpus ja-fr. Nous extrayons ces corpus d'Europarl (Koehn, 2005).

Nous nous dotons ensuite de corpus de taille moyenne ($\approx$ 400-900K phrases) pour entraîner des modèles de la langue source vers la langue pivot. Pour le japonais-anglais, nous avons rassemblé l'ensemble des corpus librement disponibles. Les thématiques sont très variées. Les corpus $\{al, es\}$-en sont à nouveau extraits d'Europarl. Enfin nous nous dotons d'un grand corpus (14 millions de phrases) pour la traduction de la langue pivot vers la langue cible. Ce corpus intègre Europarl, ce qui permet d'augmenter le vocabulaire commun avec les corpus $\{al, es\}$-en de langues européennes. Au final presque 100% du vocabulaire des corpus $\{al, es\}$-en est présent dans le corpus en-fr contre seulement 40% pour le corpus ja-en.

Pour éviter les biais causés par des structures phrastiques trop différentes d'un corpus à l'autre, nous limitons les observations à des phrases structurellement «standards». L'autre intérêt de cette restriction, c'est que les segments du corpus test (PUD, voir plus loin) sont majoritairement des

Langues	Corpus	nb phrases.	content	voc. src	voc. tgt
ja-fr	rbjafr-191204	51 659	varia	32K	27K
de-fr	rbeurdefr-1.0	52 175	Europarl	28 424	43 332
es-fr	rbeuresfr-1.0	52 807	Europarl	29 706	33 149
ja-en	rbjaen-191023	752 518	varia	150 004	137 415
de-en	rbeurdeen-1.0	973 557	Europarl	150002	56447
es-en	rbeuresen-1.0	491 387	Europarl	49144	86038
en-fr	rbenfr-192710	14 048 795	varia*	170004	170002
*ja-en***	rbjaen-191023	idem	idem	150004	137415
*es-en***	rbeuresen-1.0	idem	idem	49144	86038
*en-fr***	rbenfr-191124	14 048 795	varia*	150004	150002

TABLE 1: Corpus utilisés (* inclu Europarl ; ** pour entraîner les modèles améliorés)

.

phrases «standards». Cette sélection permet donc d'harmoniser les corpus d'entraînement et le corpus d'évaluation. Nous n'utilisons que des ressources librement disponibles et de bonne qualité, sans bruit. En plus, nous retenons les phrases qui finissent par un signe de ponctuation («. ? !») et contiennent moins de 10 chiffres et moins de 20 majuscules. Ce critère simple suffit à éliminer les phrases non standards sans trop d'erreurs. Ceci a réduit considérablement le nombre de phrases exploitées par rapport au nombre initialement disponibles, en particulier pour l'anglais-français.

Les techniques de pré et post-traitement des corpus varient d'une paire de langues à l'autre et peuvent biaiser les comparaisons. Pour uniformiser le traitement, nous utilisons un pré et post traitement a minima, avant tout pensés pour réduire la taille du vocabulaire. Les corpus ont été prétraités simplement en ajoutant une marque de début de phrase à chaque phrase, et en transformant les majuscules en minuscule en tête de phrase. Pour limiter la taille du vocabulaire et en particulier le nombre de mots inconnus, nous avons décomposé et balisé les numéraux et abréviations ("123" → "<num> 1 2 3 </num>"). Bien qu'elle permette de sensiblement réduire la taille du vocabulaire, nous n'avons pas utilisé de segmentation de type BPE (Sennrich *et al.*, 2016). Nous avons en effet estimé qu'elle aurait biaisé les comparaisons en favorisant trop les paires de langues utilisant un même système d'écriture.

Pour faciliter la comparaison avec d'autres études incluant le japonais (par exemple (Sekizawa *et al.*, 2017)), nous avons segmenté le corpus japonais en utilisant l'analyseur morphologique MeCab [1] et le dictionnaire IPADIC (Asahara & Matsumoto, 2003) [2]. Les deux sont très utilisés pour ce type de travail.

Pour l'évaluation, nous avons utilisé le corpus multilingue PUD [3] (1000 phrases). Le contenu est thématiquement très varié. Pour évaluer les modèles de base (c'est-à-dire non traduits) et les comparer à l'état de l'art, nous avons aussi extrait des corpus test de 1000 phrases, par échantillonnage des corpus d'origine.

La description des corpus de base est donnée dans le tableau 1.

1. Kudou, Taku. "MeCab : Yet Another Part-of-Speech and Morphological Analyzer". taku910.github.io (in Japanese). 23/01/2018.
2. https ://ja.osdn.net/projects/ipadic/docs/ipadic-2.7.0-manual-en.pdf/en/1/ipadic-2.7.0-manual-en.pdf.pdf
3. https ://lindat.mff.cuni.cz/repository/xmlui/handle/11234/1-2184

2.2 Méthode

La procédure est la suivante (voir aussi le Fig.1) . On construit les modèles de base $m(src, piv^1)$ et $m(piv^2, tgt)$ pour procéder à la traduction séquentielle. Puis on traduit le corpus piv^1 en tgt' à l'aide de $m(piv^2, tgt)$. On évalue alors la qualité de la traduction directe effectuée à l'aide du modèle traduit $m(src, tgt')$.

Pour tenir compte de l'influence de la qualité de la traduction sur les performances des deux techniques de traduction par pivot, nous avons réalisé les expériences à deux reprises avec un même système de traduction, mais des réglages différents. Avec le premier réglage, l'entraînement réclamait moins de temps de calcul mais la qualité des traductions de base était globalement plus modeste. Nous avons refait l'expérience en adoptant des réglages plus gourmants en puissance de calcul, mais produisant des traductions de base de meilleure qualité. Puisque l'allemand a toujours produit des résultats intermédiaires avec la première configuration, nous sommes parti du principe qu'il en serait de même avec la seconde. En conséquence, nous ne l'avons pas observé pour les traductions de qualité améliorée.

Le système de traduction neuronale utilisé est OpenNmt-py (Klein *et al.*, 2017). Pour la configuration «non améliorés», nous avons utilisé les réglages par défaut : réseau neuronal bi-récurrent, 2 couches, RNN de taille 500, word embedding de taille 500, 10 époques, beam de taille 5 pour la traduction. Cette configuration s'est révélée inadaptée pour traiter les corpus de grandes tailles. Pour améliorer les performances des modèles, la configuration «améliorée» différait sur les points suivant : 4 couches, RNN de taille 1000, word embedding de taille 600. En plus, les paramètres du traducteur ont été modifiés : batch de taille 20, beam de taille 8, longueur maximum 150. Les mots inconnus sont remplacés par les mots source qui ont un poids d'attention plus grand. Cette stratégie avantage les langues qui ont un même système d'écriture, et a fortiori un vocabulaire commun.

Nous avons utilisé BLEU (Papineni *et al.*, 2002) pour l'évaluation [4]. Nous avons aussi calculé le score Meteor (Denkowski & Lavie, 2014) mais il n'a pas produit d'information complémentaire. En conséquence, nous ne l'évoquons pas ici. On a aussi observé la fréquence des mots inconnus.

3 Résultats et discussion

La qualité des traductions fournies par les systèmes de base (voir Table 2) sont conformes aux attentes. La traduction directe avec les petits corpus obtient un score très bas. Malgré tout, BLEU est corrélé à la proximité linguistique. Plus les langues sont linguistiquement proches, meilleure est la traduction. Le fait que la qualité des traductions de PUD soit moins bonne que la traduction des corpus tests s'explique par le fait que ce corpus n'a pas de rapport thématique avec le corpus d'entraînement (et donc un faible volume de vocabulaire commun). Pour les traductions impliquant l'anglais, les scores sont inférieurs à l'état de l'art (BLEU de plus de 41.5 pour l'anglais-français dans (Shaw *et al.*, 2018), plus de 27 pour le japonais-anglais (Cromieres *et al.*, 2017) etc). Ces résultats s'expliquent par l'absence d'optimisation tant du côté de la préparation de nos corpus que du côté du réglage du système de traduction (entraînement et traduction). Nous avons certes amélioré la qualité des traductions avec la seconde configuration, mais sans pour autant chercher à rivaliser avec les meilleurs systèmes. En effet, il ne s'agissait pas dans ce travail de surpasser l'état de l'art mais de comparer

4. Nous avons utilisé multi-bleu.perl avec les réglages par défaut (https ://github.com/moses-smt/mosesdecoder/blob/master/scripts/generic/multi-bleu.perl)

	Test		PUD		Etat de l'art	
Langues	BLEU	% unk.	BLEU	% unk.	BLEU	source
ja-fr	8,63	0	2,22	0	10,03	(Blin, 2018b)
de-fr	12,86	0	4,47	0	28,63	(Bougares *et al.*, 2019)
es-fr	19,77	0	8,36	0	32,7	(Klein *et al.*, 2017)
ja-en	21,68	0.61	10,16	2.73	27,66	(Cromieres *et al.*, 2017)
de-en	29,64	0	16,15	0	42,5	(Popovic, 2017)
es-en	37,00	0	23,53	0	36,05	(Duma & Menzel, 2018)
en-fr	24,90	0.91	22,15	2.65	41,5	(Shaw *et al.*, 2018)
*ja-en***	23,18	0	10,92	0		
*es-en***	38,00	0	24,97	0		
*en-fr***	33,24	0	25,86	0		

TABLE 2: Evaluation des modèles de base (** configurarion améliorée) ; Comparaison à l'état de l'art tous corpus confondus (et à l'exclusion du corpus PUD).

deux techniques pour des niveaux égaux de préparation des corpus et de performances des modèles.

Sans surprise, quelle que soit la qualité des traductions, la traduction par pivot (voir Tables 3 et 4) obtient de meilleurs scores que la traduction directe avec corpus de base. Nous attribuons ce résultat à la différence de taille des corpus puisque le corpus pour la traduction directe est au moins 8 fois plus petit que tous les autres corpus). Le gain est plus important pour les langues européennes (proximité et corpus en commun plus grand). Mais les résultats restent toutefois faibles. (voir Fig.2)

Source	BLEU	%<unk>	gain
ja	6,74	3.33	4.52
de	10,26	0.38	5.79
es	13,49	1.38	5.13

TABLE 3: Traduction séquentielle du corpus PUD (voir Table 1 pour les traductions source→pivot correspondantes) ; gain (nombre de points) par rapport à la traduction directe ; configuration non améliorée.

Source	BLEU	%<unk>	gain
ja	6,29	5.18	4,07
de	10,05	4.42	5,58
es	14,00	5.32	5,64

TABLE 4: Evaluation des «modèles traduits» ; gain (en nombre de points) par rapport à la traduction directe ; configuration non améliorée.

Comparons les deux méthodes de traduction avec langue pivot (voir Tables 3 et 4, Fig.2). Avec une traduction de qualité modeste, la méthode séquentielle et la méthode par modèle traduit sont très proches. Par contre, on observe une différence au niveau des mots inconnus, plus nombreux avec la traduction par modèle traduit.

Si l'on augmente la qualité des traductions, les deux techniques n'ont plus la même efficacité (Table 5 et Fig.2). La méthode par modèle traduit est meilleure ($\approx +3$). Ce constat est le même quelle que soit la proximité linguistique entre les langues source et la langue pivot, et quelle que soit la quantité de vocabulaire commun entre les deux corpus piv^1 et piv^2. Par extrapolation, nous pouvons avancer que ce résultat vaut quelle que soit la distance thématique entre le corpus d'entraînement et le corpus d'évaluation. La qualité des traductions est malheureusement trop basse pour faire une comparaison qualitative pertinente. On observe des erreurs dans tous les domaines, lexico, syntaxiques et logiques.

	Source	en	fr
Traduction séquentielle	ja	10,92	8,38
	es	24.79	16.76
Model traduit	ja	_	10,11
	es	_	19,79

TABLE 5: Score BLEU pour la traduction séquentielle et par «modèle traduit», configuration améliorée

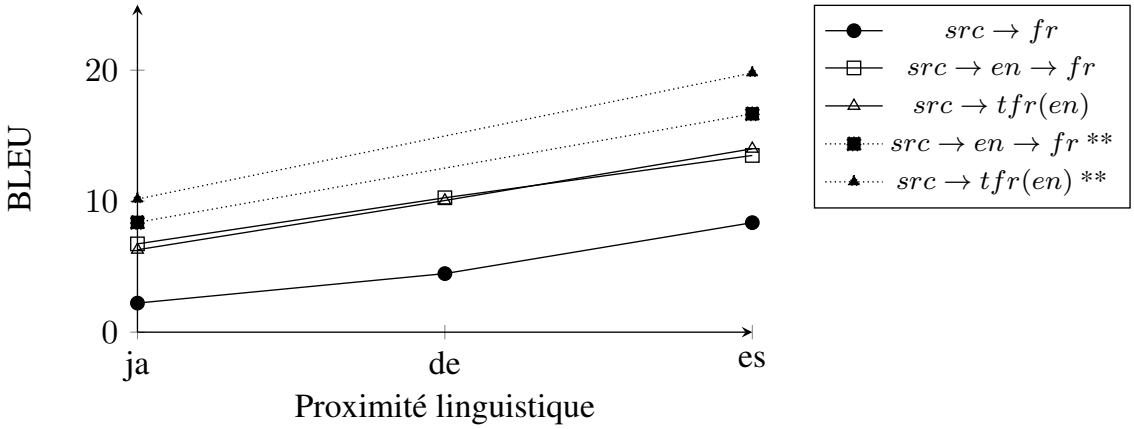

FIGURE 2: Traduction séquentielle et par modèle traduit (** Configuration améliorée)

4 Conclusion

Nos résultats montrent qu'utiliser un «modèle traduit» produit de meilleurs résultats qu'une traduction séquentielle par pivot, sous réserve que les modèles de traduction source→pivot et pivot→cible soient de bonne qualité. Sinon, la traduction séquentielle reste compétitive. Ces résultats confirment les observations de (Currey & Heafield, 2019). Ils montrent en plus que les résultats vont dans le même sens quelles que soient la proximité entre les langues source et cible et la quantité de vocabulaire commun entre les corpus. Enfin, ils montrent que la supériorité de la traduction par «modèle traduit» est vraie même sans recourir à un corpus multilingue.

Remerciements

Nous remercions chaleureusement le Centre de Calcul CNRS / IN2P3 (Lyon - France) pour la mise à disposition des moyens informatiques nécessaires à ces travaux.

Références

AHARONI R., JOHNSON M. & FIRAT O. (2019). Massively Multilingual Neural Machine Translation. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies, Volume 1 (Long and Short Pa-*

pers), p. 3874–3884, Minneapolis, Minnesota : Association for Computational Linguistics. DOI : 10.18653/v1/N19-1388.

ARIVAZHAGAN N., BAPNA A., FIRAT O., LEPIKHIN D., JOHNSON M., KRIKUN M., CHEN M. X., CAO Y., FOSTER G., CHERRY C., MACHEREY W., CHEN Z. & WU Y. (2019). Massively Multilingual Neural Machine Translation in the Wild : Findings and Challenges. arXiv : 1907.05019.

ASAHARA M. & MATSUMOTO Y. (2003). Ipadic user manual. ipadic-2.7.0-manual-en.pdf.pdf.

BLIN R. (2018a). Automatic evaluation of alignments without using a gold-corpus - example with french-japanese aligned corpora. In S. KIYOAKI, Éd., *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Paris, France : European Language Resources Association (ELRA). 23_W29.pdf.

BLIN R. (2018b). Traduction automatique du japonais vers le français : Bilan et perspectives. In *Traitement Automatique du Langage Naturel*, Rennes, France. HAL : hal-01796313.

BOUGARES F., WOTTAWA J., BAILLOT A., BARRAULT L. & BARDET A. (2019). LIUM's Contributions to the WMT2019 News Translation Task : Data and Systems for German-French Language Pairs. In *Proceedings of the Fourth Conference on Machine Translation (Volume 2 : Shared Task Papers, Day 1)*, p. 129–133, Florence, Italy : Association for Computational Linguistics.

CROMIERES F., DABRE R., NAKAZAWA T. & KUROHASHI S. (2017). Kyoto university participation to wat 2017. In *Proceedings of the 4th Workshop on Asian Translation (WAT2017)*, p. 146–153, Taipei, Taiwan : Asian Federation of Natural Language Processing.

CURREY A. & HEAFIELD K. (2019). Zero-resource neural machine translation with monolingual pivot data. In *Proceedings of the 3rd Workshop on Neural Generation and Translation*, p. 99–107, Hong Kong : Association for Computational Linguistics. DOI : 10.18653/v1/D19-5610.

DENKOWSKI M. & LAVIE A. (2014). Meteor Universal : Language Specific Translation Evaluation for Any Target Language. In *Proceedings of the EACL 2014 Workshop on Statistical Machine Translation*.

DUMA M.-S. & MENZEL W. (2018). Translation of Biomedical Documents with Focus on Spanish-English. In *Proceedings of the Third Conference on Machine Translation : Shared Task Papers*, p. 637–643, Belgium, Brussels : Association for Computational Linguistics. DOI : 10.18653/v1/W18-6444.

KLEIN G., KIM Y., DENG Y., SENELLART J. & RUSH A. (2017). OpenNMT : Open-Source Toolkit for Neural Machine Translation. In *Proceedings of ACL 2017, System Demonstrations*, p. 67–72, Vancouver, Canada : Association for Computational Linguistics.

KOCMI T. & BOJAR O. (2019). Transfer Learning across Languages from Someone Else's NMT Model. arXiv : 1909.10955.

KOEHN P. (2005). Europarl : A Parallel Corpus for Statistical Machine Translation. In *Proceedings of Machine Translation Summit X*, p. 79–86.

KOEHN P. & KNOWLES R. (2017). Six Challenges for Neural Machine Translation. In *Proceedings of the First Workshop on Neural Machine Translation*, p. 28–39, Vancouver : Association for Computational Linguistics. DOI : 10.18653/v1/W17-3204.

PAPINENI K., ROUKOS S., WARD T. & ZHU W.-J. (2002). BLEU : a Method for Automatic Evaluation of Machine Translation. In *Proc. ACL*, p. 311–318.

POPOVIC M. (2017). Comparing Language Related Issues for NMT and PBMT between German and English. *The Prague Bulletin of Mathematical Linguistics*, **108**. DOI : 10.1515/pralin-2017-0021.

RIKTERS M., PINNIS M. & KRIŠLAUKS R. (2018). Training and Adapting Multilingual NMT for Less-resourced and Morphologically Rich Languages. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan : European Language Resources Association (ELRA).

SEKIZAWA Y., KAJIWARA T. & KOMACHI M. (2017). Improving Japanese-to-English Neural Machine Translation by Paraphrasing the Target Language. In *Proceedings of the 4th Workshop on Asian Translation (WAT2017)*, p. 64–69, Taipei, Taiwan : Asian Federation of Natural Language Processing.

SENNRICH R., HADDOW B. & BIRCH A. (2016). Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1 : Long Papers)*, p. 1715–1725, Berlin, Germany : Association for Computational Linguistics. DOI : 10.18653/v1/P16-1162.

SHAW P., USZKOREIT J. & VASWANI A. (2018). Self-Attention with Relative Position Representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies, Volume 2 (Short Papers)*, p. 464–468, New Orleans, Louisiana : Association for Computational Linguistics. DOI : 10.18653/v1/N18-2074.