

# Adaptation incrémentale de modèles de traduction neuronaux

Christophe Servan   Josep Crego   Jean Senellart

SYSTRAN, 5 Rue Feydeau, 75002 Paris, France  
prenom.nom@systrangroup.com

## RÉSUMÉ

---

L'adaptation au domaine est un verrou scientifique en traduction automatique. Il englobe généralement l'adaptation de la terminologie et du style, en particulier pour la post-édition humaine dans le cadre d'une traduction assistée par ordinateur. Avec la traduction automatique neuronale, nous étudions une nouvelle approche d'adaptation au domaine que nous appelons "spécialisation" et qui présente des résultats prometteurs tant dans la vitesse d'apprentissage que dans les scores de traduction. Dans cet article, nous proposons d'explorer cette approche.

## ABSTRACT

---

### **Incremental adaptation of neural machine translation models**

Domain adaptation is a key feature in Machine Translation. It generally encompasses terminology and style adaptation, especially for human post-editing workflows in Computer Assisted Translation (CAT). With Neural Machine Translation (NMT), we study a new approach of domain adaptation that we call "specialization" and which is showing promising results both in the learning speed and in adaptation accuracy. In this paper, we propose to explore this approach.

---

**MOTS-CLÉS :** Traduction Automatique Neuronale, Adaptation au Domaine, étude préliminaire.

**KEYWORDS:** Neural Machine Translation, Domain Adaptation, preliminary study.

---

## 1 Introduction

Suite aux avancées de la traduction automatique neuronale (TAN) ces dernières années (Kalchbrenner & Blunsom, 2013; Sutskever *et al.*, 2014; Bahdanau *et al.*, 2014), nous observons que les performances obtenues dans les campagnes d'évaluations sont comparables, voire meilleures, que les approches de traduction automatique statistiques (TAS) (Cettolo *et al.*, 2015; Bojar *et al.*, 2015, 2016). Ces succès obtenus nous permettent d'explorer des problématiques qui se posaient déjà avec la TAS. L'une de ces problématiques concerne l'adaptation au domaine.

Malgré la quantité de données disponibles, ces dernières sont généralement hors-domaine. L'idée sous-jacente est de mieux utiliser les données spécifiques à un domaine, conjointement aux données hors-domaine. Intuitivement, la première idée consiste à combiner les données hors-domaine à celles du domaine. Or, outre la question de la durée d'apprentissage (qui peut durer jusqu'à plusieurs semaines), se pose surtout celle de la représentativité des données du domaine par rapport aux données hors-domaine.

**Contributions** Cet article propose une méthode permettant d'améliorer la représentativité des données du domaine par rapport aux données hors domaine en se basant sur les précédents travaux de réentraînement initialement proposé par (Luong & Manning, 2015). Plus généralement, nous présentons une étude de cette approche d'adaptation incrémentale s'insérant dans le cadre expérimental de la traduction assistée par ordinateur.

Les expériences sont tenues compte d'un domaine spécialisé : des transcriptions de présentations (issues de TED). La tâche de traduction du cadre expérimental sont de l'anglais vers le français et de l'anglais vers l'allemand. Les données sont issues de campagnes d'évaluations ou de dépôts libres de droits et seront mis à disposition afin d'assurer la reproductibilité des résultats.

Cet article débute par une étude de l'état de l'Art, suivi de la description de l'approche. Dans la section 4 nous présentons notre protocole d'évaluation, contenant une étude des résultats et une discussion autour de l'approche. Enfin nous concluons cet article.

## 2 État de l'Art

Ces dernières années, les techniques d'adaptation au domaine pour la traduction ont été particulièrement étudiées. Ces approches peuvent être appliquées à trois différents niveaux : avant, pendant ou après la phase d'apprentissage initiale. Dans le cadre d'une traduction assistée par ordinateur (TAO), la plupart des approches sont des pré-traitements ou des post-traitements de la phase d'apprentissage initiale.

Ainsi, les pré-traitements concernent généralement les données d'apprentissage. Citons en exemple la sélection de données proposée par Lü *et al.* (2007) et améliorée par Gao & Zhang (2002) et bien d'autres (Moore & Lewis, 2010; Axelrod *et al.*, 2011). Elle est connue pour être efficace (Lambert *et al.*, 2011; Cettolo *et al.*, 2014) malgré son principal défaut : les mots et séquences de mots rares ne sont pas identifiés et sont donc rejetés (Wuebker *et al.*, 2014). En cela, l'approche permettant l'extraction de séquences de mots rares (Gascó *et al.*, 2012) palie en partie à ce problème (Wuebker *et al.*, 2014). Cependant, ce type d'approche ne nous permet pas d'adapter un modèle de traduction générique déjà entraîné.

Lors de l'apprentissage, il existe des approches permettant de pondérer des données d'apprentissage au niveau du corpus (du document) ou de la phrase (Shah *et al.*, 2010). Mais le principal inconvénient de ce type de pré-traitement est la nécessité de relancer la phase d'apprentissage pour que celle-ci soit efficace. Or, pour des besoins pratiques, cette étude focalise sur la possibilité de profiter d'un modèle déjà entraîné afin de l'adapter à un domaine particulier. Cependant, l'idée de caractériser les données d'apprentissage selon un domaine particulier, n'est pas nouvelle. Cette idée se rapproche des travaux proposés par Sennrich *et al.* (2016a) qui visent à adapter le modèle de traduction en fonction du niveau de politesse dans la phrase. Dans ces travaux ils utilisent un marqueur de niveau de langage, ce qui rend toute adaptation postérieure à l'apprentissage d'un nouveau domaine compliquée. Enfin, Zhang *et al.* (2016) proposent l'utilisation de marqueur afin de caractériser les thématiques de traduction.

L'approche, dite de « spécialisation », étudiée ici, appartient à la famille de techniques de post-traitement qui ont pour but la mise-à-jour ou l'optimisation de paramètres d'un modèle de traduction existant. Elle est dérivée de celle proposée par Doulamis *et al.* (2000) pour le traitement de l'image et appliquée à la traduction une première fois avec succès par Luong & Manning (2015). Ces derniers proposent d'entraîner un modèle puis de ré-entraîner ce même modèle sur un corpus entier spécifique à un domaine particulier à travers une douzaine de cycles supplémentaires. Cette approche paraît

adaptée à la problématique d'une mise-à-jour rapide d'un modèle. A notre connaissance, il n'existe pas de travaux qui ont étudié l'impact de cette approche et son comportement. Nous proposons donc d'étudier cette approche dans le cadre d'une mise-à-jour rapide du modèle en faisant varier les quantités de données.

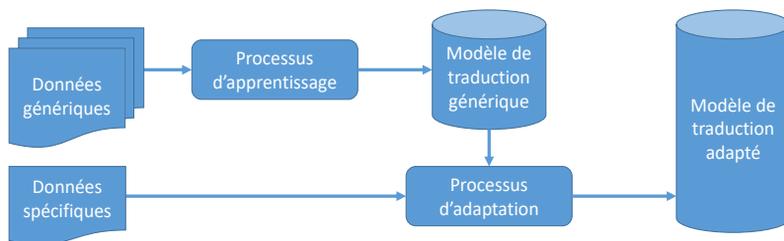


FIGURE 1 – Le modèle de traduction générique est appris sur les données génériques, puis, ce modèle générique est ré-entraîné avec des données spécifiques au domaine afin d'obtenir un modèle adapté au domaine (ou à la tâche).

### 3 Approche

L'approche de spécialisation vise à adapter de manière directe ou incrémentale un modèle de traduction neuronal générique en s'inspirant des travaux sur l'adaptation incrémentale en traduction automatique statistique (Blain *et al.*, 2012; Cettolo *et al.*, 2014). Ces approches s'insèrent dans un cadre applicatif particulier : la traduction assistée par ordinateur. Dans ce cadre, le traducteur corrige ou post-édite les sorties d'un système de traduction automatique. Ces travaux ont montré l'intérêt d'ajouter de manière incrémentale de nouvelles informations dans un modèle de traduction fondé sur les paires de segments. Ces informations peuvent être d'ordre lexical ou stylistique, ce qui peut déboucher éventuellement sur une adaptation au style du traducteur.

L'idée principale de l'approche est donc de spécialiser un modèle générique déjà entraîné sur des données génériques (i.e. : venant de plusieurs domaines différents). La figure 1 présente le cycle d'adaptation que nous utilisons lors de nos expériences.

Lors de ce processus de spécialisation, sont réestimées les probabilités conditionnelles  $p(y_1, \dots, y_m | x_1, \dots, x_n)$  où  $(x_1, \dots, x_n)$  est la séquence d'entrée de taille  $n$  et  $(y_1, \dots, y_m)$  est la séquence de sortie correspondante de taille  $m$ . L'adaptation a donc comme point de départ l'état du réseau de neurones, d'une manière similaire à une pré-configuration des paramètres du modèle.

Le résultat est un modèle adapté (ou spécialisé) à une tâche ou un domaine particulier.

### 4 Expériences

Ne disposant pas de données post-éditées et afin d'être le plus reproductible possible, nous proposons d'étudier l'approche de spécialisation sur les données issues de l'évaluation IWSLT 2015 (Cettolo *et al.*, 2015), utilisant les traductions de transcriptions de vidéos de TED. Les paires de langues concernées sont l'anglais vers le français et l'anglais vers l'allemand. La mesure de comparaison

Corpus	Nbr lignes	Nbr mots (en)	Nbr mots (fr)	Nbr lignes	Nbr mots (en)	Nbr mots (de)
BOOKS	120 K	2,9 M	3,0 M	46,1 K	1,1 M	1,1 M
Europarl V.7	2 M	55,6 M	64,1 M	1,9 M	53,6 M	51,0 M
IT	371 K	4,3 M	5,3 M	218 K	2,7 M	2,7 M
JRC	603 K	16,2 M	18,4 M	601 K	16,8 M	15,2 M
News-Commentary V.11	203 K	5,1 M	6,4 M	220 K	5,5 M	5,6 M
Total Générique	3,3 M	84,3 M	97,4 M	3,0 M	80 M	75,8 M
WIT <sup>3</sup>	217,9 K	4,5 M	4,9 M	192,7 K	4,0 M	3,6 M
IWSLT2014 (test)	1305	25 K	29 K	1305	23 K	25 K

TABLE 1 – Détail des données d’apprentissage pour les tâches anglais-français et anglais-allemand. Les cinq premiers corpus sont associés au domaine générique. Le corpus WIT<sup>3</sup> est le corpus du domaine spécifique sur lequel nous souhaitons adapter le modèle générique. Le corpus de test est celui de l’évaluation IWSLT2014. Les notations K et M correspondent respectivement à  $10^3$  et  $10^6$ .

utilisée dans cette étude est le score BLEU (Papineni *et al.*, 2002), bien que controversée, elle reste la mesure étalon lors des dernières campagnes d’évaluation (Bojar *et al.*, 2016; Cettolo *et al.*, 2016).

L’approche de traduction automatique neuronale (TAN) utilisée dans cet article est fondée sur une approche séquence-vers-séquence (Sutskever *et al.*, 2014) combinée à une approche d’attention (Luong *et al.*, 2015), à travers l’outil *OpenNMT* (Klein *et al.*, 2017).

Les données sont pré-traitées en utilisant une méthode de découpage en sous-mots issue de l’algorithme de compression de données : « *byte pair encoding* » (Sennrich *et al.*, 2016b) avec 30 000 opérations. Nous avons conservé les 32 000 mots plus fréquents pour les langues source et cible. Le modèle utilise une architecture récurrente (bi-LSTM) de 4 couches cachées avec 800 cellules chacune. Lors de l’apprentissage, le mini-batch est fixé à 64 avec une régularisation de type « dropout » fixée à 0,3. Les modèles génériques sont appris sur 18 cycles avec un taux d’apprentissage à 1,0 auquel il est appliqué une dégradation de facteur 0,7 à partir du cycle 10. Le système générique, utilisant ces paramètres, a mis plus de 12 jours à être appris avec une carte Nvidia GeForce GTX 1080.

## 4.1 Données

Les données utilisées sont celles mises à disposition pour l’évaluation WMT 16 (Bojar *et al.*, 2016) pour Europarl et News Commentary. Les données « IT » viennent de la fusion des corpus Ubuntu, KDE4 et KDE-doc disponibles sur le site de OPUS<sup>1</sup> (Tiedemann, 2012). De même, les corpus JRC et BOOKS sont également issus de cette même source. Le corpus WIT<sup>3</sup> (Cettolo *et al.*, 2012) est une collection de transcriptions de vidéos de TED avec leurs traductions. Ces données sont régulièrement utilisées pour les évaluations IWSLT d’où sont issues nos données de test (IWSLT2014). L’ensemble des corpus utilisé est détaillé dans le tableau 1.

## 4.2 Résultats

Le tableau 2 présente l’ensemble des résultats obtenus lors de nos expériences. Nous proposons donc d’étudier le comportement des modèles réentraînés avec 5 cycles d’apprentissage supplémentaires en faisant varier les quantités de données de réentraînement. Ainsi, nous avons découpé le corpus WIT<sup>3</sup> selon cinq tailles différentes : 500, 1 000, 5 000, 10 000 lignes et enfin le corpus entier. Ceci afin de mesurer l’impact de ces quantités de données sur les résultats.

1. <http://opus.lingfil.uu.se>

Modèles	<i>en-fr</i>	<i>en-de</i>	temps d'apprentissage
<i>générique</i>	33,82	20,07	12 jours
<i>spécialisé (WIT<sup>3</sup>-0,5K)</i>	33,26	19,00	5 min
<i>spécialisé (WIT<sup>3</sup>-1K)</i>	34,87	20,30	8 min
<i>spécialisé (WIT<sup>3</sup>-5K)</i>	36,95	21,51	15 min
<i>spécialisé (WIT<sup>3</sup>-10K)</i>	37,42	22,38	17 min
<i>spécialisé (WIT<sup>3</sup>-entier)</i>	<b>39,53</b>	<b>25,53</b>	168 min ( $\approx$ 2h45)
<i>générique+WIT<sup>3</sup>-entier</i>	35,60	22,68	12 jours

TABLE 2 – Scores de BLEU pour la spécialisation avec différentes tailles du corpus spécifique *WIT<sup>3</sup>*. A titre indicatif, les temps d'apprentissage et de spécialisation sont indiqués dans la dernière colonne.

Nous observons ici deux systèmes de base, le premier est un système (noté *générique*) appris sur les données que nous avons classées comme génériques, puis un second système appris avec ces mêmes données génériques que nous avons concaténés avec les données spécialisées (noté *générique+WIT<sup>3</sup>*). Le second obtient de meilleurs résultats sur notre corpus de test IWSLT2014 aussi bien en français qu'en allemand. A partir du premier modèle, nous proposons donc d'appliquer l'approche de spécialisation décrite dans la partie 3 en utilisant les corpus de tailles différentes.

Ainsi, la spécialisation avec 500 lignes n'apporte rien, au contraire, les scores de BLEU dans les deux langues décroissent : -0,6 et -1 point respectivement pour l'anglais vers le français et l'allemand. On observe une première progression des scores lorsqu'on spécialise le modèle avec 1000 lignes. Dans la tâche de traduction vers le français, le modèle spécialisé améliore de 1 point de BLEU. Dans la tâche vers l'allemand, la progression est plus timide avec +0,3 point de BLEU. Les améliorations sont plus marquées lorsqu'on spécialise avec 5000 lignes pour les deux tâches. Ainsi, les résultats passent de 33,82 à 36,87 pour le français et de 20,07 à 21,51 pour l'allemand. La progression est toujours présente avec une quantité de corpus supplémentaire (10 000 lignes) les scores sont encore améliorés de 0,5 point pour le français et de 0,8 point pour l'allemand. Enfin, les meilleurs scores sont obtenus avec l'utilisation du corpus en entier, où le score du modèle générique français est amélioré de presque 6 points BLEU et l'allemand de 5,5 points BLEU.

Lorsque nous comparons ces résultats de la spécialisation avec les résultats obtenus avec le modèle incluant l'ensemble des données de *WIT<sup>3</sup>*, nous observons que ce dernier est dépassé en termes de score dès lors que nous ajoutons 5 000 lignes, soit moins de 3 pour cent du corpus initial pour les deux paires de langue.

### 4.3 Discussion

Source :	So your 20s are the time to educate yourself about your body ...
Reference :	Donc la vingtaine est le moment de s' informer sur son corps ...
<i>générique</i>	Donc , vos <b>années 20 sont l' occasion</b> de vous éduquer sur votre corps ...
<i>spécialisé (WIT<sup>3</sup>-0.5K)</i>	Donc , vos <b>20 s</b> sont le moment d' <b>apprendre à votre corps</b> ...
<i>spécialisé (WIT<sup>3</sup>-1K)</i>	Donc vos <b>20èmes</b> sont le moment d' être éduqués sur votre corps ...
<i>spécialisé (WIT<sup>3</sup>-5K)</i>	Donc , vos <b>20 ans</b> sont le moment d' apprendre votre corps ...
<i>spécialisé (WIT<sup>3</sup>-10K)</i>	Donc , vos <b>20 ans</b> sont le moment d' éduquer votre corps ...
<i>spécialisé (WIT<sup>3</sup>-entier)</i>	Donc vos <b>20 ans</b> est le moment d' éduquer votre corps ...
<i>générique+WIT<sup>3</sup>-entier</i>	Donc vos <b>20s</b> sont le moment de vous éduquer sur votre corps ...

TABLE 3 – Exemples de traductions correspondant à différents points. Le rouge, bleu et vert sont des traductions notées respectivement *mauvaise*, *acceptable* et *bonne* .

Malgré les performances obtenues avec cette approche, il est nécessaire d'avoir une quantité minimale de corpus. Ainsi, l'idée d'utiliser quelques phrases ou un paragraphe afin de spécialiser le modèle ne semble pas suffisant. Au contraire, utiliser une si petite quantité de données semble perturber le

modèle au point de donner de moins bons résultats que si l'on ne fait rien (utilisation du modèle générique). Reste cependant l'idée de réutiliser des données du corpus d'apprentissage en plus des données spécifiques, mais nous n'avons pas eu le temps de faire ce type d'expériences.

Le tableau 3 présente un exemple tiré du corpus de test. Nous avons pris un exemple dans les sorties qui illustre bien la progression de la spécialisation ou l'adaptation des modèles à la tâche de traduction. Ainsi, nous pouvons voir les améliorations dues aux différentes quantités de données qui ont été ajoutées. Même si les sorties ne correspondent pas exactement à la référence, nous considérons la traduction comme bonne selon des critères syntaxiques et sémantiques.

De plus, un point particulier à mettre en avant est le gain de temps que permet cette approche. Outre le fait des performances, cette approche permet de s'affranchir de l'obligation de réentraîner depuis le début un modèle de traduction adapté à une tâche ou un domaine particulier. Lorsqu'on se met dans un cadre particulier de la traduction assistée par ordinateur, ce cadre nécessite une adaptation à un domaine particulier en un temps relativement court. De fait, cette approche autorise une mise-à-jour journalière du modèle de traduction et répond à cette contrainte. En effet, il faut moins de vingt minutes (pour traiter 5 000 et 10 000 lignes, voir le tableau 2) afin d'adapter et de spécialiser un modèle générique et obtenir des performances intéressantes. Le gain de temps et d'utilisabilité du modèle ne sont que plus intéressants.

Enfin, nous remarquons que cette approche pourrait s'apparenter à une méthode d'optimisation du modèle, telle que présentée par Och (2003) pour les modèles de traduction à base de paires de segments. Cette phase de spécialisation pourrait, à l'avenir, se focaliser sur une métrique particulière telle que le BLEU, le TER ou encore le METEOR.

## 5 Conclusion

Dans cet article nous avons proposé une courte étude sur une approche d'adaptation au domaine des modèles de traduction neuronaux : la spécialisation. Cette approche d'adaptation propose des améliorations significatives en termes de scores BLEU qui sont perceptibles dans les sorties. Ces gains permettent de s'affranchir d'un entraînement long et coûteux (plusieurs jours, voire plusieurs semaines contre moins de vingt minutes) afin d'obtenir des performances significatives.

Dans le cadre d'un processus de traduction assistée par ordinateur, ce processus de spécialisation pour les modèles de traduction neuronaux semble particulièrement pertinent, permettant une rapide mise-à-jour des modèles. L'une des perspectives de ces premiers travaux est d'étudier et de mesurer, dans un cadre *in vivo*, l'impact de cette approche et vérifier quels seraient les gains obtenus du point de vue des traducteurs. Toujours dans ce cadre, nous souhaitons étudier l'impact d'une adaptation incrémentale sur les performances de traduction.

Enfin cette approche de spécialisation ouvre l'étude à d'autres approches classiques pour l'adaptation au domaine telles que la pondération de corpus ou de phrase lors de l'apprentissage (Shah *et al.*, 2010, 2011). Nous pourrions également explorer l'utilisation de marqueurs de domaine tels qu'ils ont été introduit par Sennrich *et al.* (2016a); Zhang *et al.* (2016), ce qui permettrait d'améliorer la représentativité des données dans le cadre de ce type d'approche. Enfin, cette phase de spécialisation pourrait se focaliser sur une métrique automatique afin d'optimiser le modèle selon cette même métrique.

# Références

- AXELROD A., HE X. & GAO J. (2011). Domain adaptation via pseudo in-domain data selection. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, p. 355–362, Edinburgh, Scotland, UK. : Association for Computational Linguistics.
- BAHDANAU D., CHO K. & BENGIO Y. (2014). Neural machine translation by jointly learning to align and translate. *CoRR*, **abs/1409.0473**. Demoeo at NIPS 2014 : <http://lisa.iro.umontreal.ca/mt-demo/>.
- BLAIN F., SCHWENK H. & SENELLART J. (2012). Incremental adaptation using translation information and post-editing analysis. In *IWSLT*, p. 229–236.
- BOJAR O., CHATTERJEE R., FEDERMANN C., GRAHAM Y., HADDOW B., HUCK M., JIMENO YEPES A., KOEHN P., LOGACHEVA V., MONZ C., NEGRI M., NEVEOL A., NEVES M., POPEL M., POST M., RUBINO R., SCARTON C., SPECIA L., TURCHI M., VERSPOOR K. & ZAMPIERI M. (2016). Findings of the 2016 conference on machine translation. In *Proceedings of the First Conference on Machine Translation*, p. 131–198, Berlin, Germany.
- BOJAR O., CHATTERJEE R., FEDERMANN C., HADDOW B., HUCK M., HOKAMP C., KOEHN P., LOGACHEVA V., MONZ C., NEGRI M., POST M., SCARTON C., SPECIA L. & TURCHI M. (2015). Findings of the 2015 workshop on statistical machine translation. In *Proceedings of the Tenth Workshop on Statistical Machine Translation*, p. 1–46, Lisbon, Portugal.
- CETTOLO M., BERTOLDI N., FEDERICO M., SCHWENK H., BARRAULT L. & SERVAN C. (2014). Translation project adaptation for mt-enhanced computer assisted translation. *Machine Translation*, **28**(2), 127–150.
- CETTOLO M., GIRARDI C. & FEDERICO M. (2012). Wit<sup>3</sup> : Web inventory of transcribed and translated talks. In *Proceedings of the 16<sup>th</sup> Conference of the European Association for Machine Translation (EAMT)*, p. 261–268, Trento, Italy.
- CETTOLO M., NIEHUES J., STUKER S., BENTIVOGLI L., CATTONI R. & FEDERICO M. (2015). The iwslt 2015 evaluation campaign. In *the proceedings of the International Workshop on Spoken Language Translation (IWSLT)*.
- CETTOLO M., NIEHUES J., STUKER S., BENTIVOGLI L., CATTONI R. & FEDERICO M. (2016). The iwslt 2016 evaluation campaign. In *the proceedings of the International Workshop on Spoken Language Translation (IWSLT)*.
- DOULAMIS A. D., DOULAMIS N. D. & KOLLIAS S. D. (2000). On-line retrainable neural networks : improving the performance of neural networks in image analysis problems. *IEEE Transactions on Neural Networks*, **11**(1), 137–155.
- GAO J. & ZHANG M. (2002). Improving language model size reduction using better pruning criteria. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, p. 176–182, Stroudsburg, PA, USA : Association for Computational Linguistics.
- GASCÓ G., ROCHA M.-A., SANCHIS-TRILLES G., ANDRÉS-FERRER J. & CASACUBERTA F. (2012). Does more data always yield better translations? In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, p. 152–161.
- KALCHBRENNER N. & BLUNSOM P. (2013). Recurrent continuous translation models. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*.
- KLEIN G., KIM Y., DENG Y., SENELLART J. & RUSH A. (2017). OpenNMT : Open-Source Toolkit for Neural Machine Translation. *ArXiv e-prints*. 1701.02810.

- LAMBERT P., SCHWENK H., SERVAN C. & ABDUL-RAUF S. (2011). Investigations on translation model adaptation using monolingual data. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, p. 284–293, Edinburgh, Scotland : Association for Computational Linguistics.
- LÜ Y., HUANG J. & LIU Q. (2007). Improving statistical machine translation performance by training data selection and optimization. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, p. 343–350.
- LUONG T. & MANNING C. D. (2015). Stanford neural machine translation systems for spoken language domains. In *IWSLT2015*, Da Nang, Vietnam.
- LUONG T., PHAM H. & MANNING C. D. (2015). Effective approaches to attention-based neural machine translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, p. 1412–1421, Lisbon, Portugal.
- MOORE R. & LEWIS W. (2010). Intelligent Selection of Language Model Training Data. In *ACL (Short Papers)*, p. 220–224, Uppsala, Sweden.
- OCH F. J. (2003). Minimum error rate training in statistical machine translation. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1*, p. 160–167.
- PAPINENI K., ROUKOS S., WARD T. & ZHU W.-J. (2002). Bleu : a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, p. 311–318, Philadelphia, Pennsylvania, USA.
- SENNRICH R., HADDOW B. & BIRCH A. (2016a). Controlling politeness in neural machine translation via side constraints. In *Proceedings of the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics*.
- SENNRICH R., HADDOW B. & BIRCH A. (2016b). Improving neural machine translation models with monolingual data. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1 : Long Papers)*, p. 86–96, Berlin, Germany : Association for Computational Linguistics.
- SHAH K., BARRAULT L. & SCHWENK H. (2010). Translation model adaptation by resampling. In *Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR, WMT '10*, p. 392–399.
- SHAH K., BARRAULT L. & SCHWENK H. (2011). Parametric weighting of parallel data for statistical machine translation. In *The 5th International Joint Conference on Natural Language Processing*, Chiang Mai (Thailand).
- SUTSKEVER I., VINYALS O. & LE Q. V. (2014). Sequence to sequence learning with neural networks. In *NIPS*.
- TIEDEMANN J. (2012). Parallel data, tools and interfaces in opus. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'2012)*.
- WUEBKER J., NEY H., MARTÍNEZ-VILLARONGA A., GIMÉNEZ A., JUAN A., SERVAN C., DYMETMAN M. & MIRKIN S. (2014). Comparison of data selection techniques for the translation of video lectures. In *AMTA*.
- ZHANG J., LI L., WAY A. & LIU Q. (2016). Topic-informed neural machine translation. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics : Technical Papers*, p. 1807–1817, Osaka, Japan.