



**HAL**  
open science

## Discussion on Bilingual Cognition in International Exchange Activities

Mieradilijiang Maimaiti, Xiaohui Zou

► **To cite this version:**

Mieradilijiang Maimaiti, Xiaohui Zou. Discussion on Bilingual Cognition in International Exchange Activities. 2nd International Conference on Intelligence Science (ICIS), Nov 2018, Beijing, China. pp.167-177, 10.1007/978-3-030-01313-4\_17. hal-02118842

**HAL Id: hal-02118842**

**<https://inria.hal.science/hal-02118842v1>**

Submitted on 3 May 2019

**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.



Distributed under a Creative Commons Attribution 4.0 International License

# Discussion on Bilingual Cognition in International Exchange Activities

Mieradilijiang Maimaiti<sup>1\*</sup>[0000-0002-4231-6519] and Xiaohui Zou<sup>2\*</sup>[0000-0002-5577-8245]

<sup>1</sup> Department of Computer Science and Technology, Tsinghua University, Beijing, China  
meadljmm15@mails.tsinghua.edu.cn

<sup>2</sup> Sino-American Saelle Research Center, Beijing, China  
geneculture@icloud.com

**Abstract.** This article aims to explore the features, mechanisms, and applications of bilingual cognition in international communication activities. Our main idea is: First, clarify the mother tongue of each international exchange activities (IEAs) and prepare some prerequisites which are related to the discussion. Then, make full use of the information and intelligent network tools to bring out the subjective initiative of both parties, while conducting the corresponding research on daily terms and professional terms, and generate the two series of bilingual phrase table. Finally, use the machine translation (MT) and translation memory tools to help them make the necessary preparations or exercises. Meanwhile, we propose the novel and efficient mixed transfer learning (MTL) approach. As a result, when the two parties communicate with each other, as well as via online or off-line communicate, the kind of tacit agreement would have been created between them. If so, it will have been leveraged among them repeatedly rather than just one time and will have targeted multiple times. Its significance lies in: This process and habit of human-computer interaction will better reveal the characteristics of bilingual cognition based on this article. Experiments on low-resource datasets show that our approach is effective, significantly outperform the state-of-the-art methods and yield improvements of up to 4.13 BLEU points.

**Keywords:** Bilingual Cognition · International Exchange Activities · Human-Computer Interaction · Neural Machine Translation · Transfer Learning.

## 1 Introduction

Recent days artificial intelligence technologies made remarkable progress, many terrific ideas, various mechanisms, and marvelous applications are designed and emerged both in our daily life and study life. Meanwhile, cognitive computation [12, 11] filed has also obtained outstanding achievements in the international exchange activities (IEAs). Likewise, bilingual cognition [13] also plays a vital role during the international exchanging with each other, and many people are trying to take advantages of human-computer interaction (HCI) mechanisms or applications to accelerate their communications and improve the quality of communications by avoiding the low efficiencies.

---

\*Corresponding author.

As we are taking part in IEAs, the majority of us would feel that it is time-consuming and hard to avoid unexpected misunderstandings at a certain level, and needs to higher cost frequently. Since, if the bilingual person even fluent to speak both two languages but sometimes also need to hire some professional translators or need to buy some artificial intelligence products, such as off-line spoken translator tools, some smart mobile devices, and simultaneous translation system. Moreover, because of the quality of these portable smart mobile devices or facilities not good enough at some special occasion, even it might cause some misunderstandings unexpectedly. Although there are many demands for IEAs, however, these demands are not be solved efficiently and effectively. In addition, many useful approaches have been proposed by numbers of researchers and scientists. The methods which could be used to improve the quality of international communications, even some essential features, convenience applications, popular mechanisms and the power of bilingual cognition.

Besides, except for these strategies, numbers of artificial intelligence algorithms, marvelous ideas, terrific mechanisms and outstanding approaches have been created in the international communication field. Such as online machine translation engines, online network applications, smart mobile devices, and off-line machine translation facilities. Even if the translation system which can be used by taking photos or by recognizing the voice of users, the core technology is still machine translation. During the IEAs, we need to use high-resource languages (HRLs) and low-resource languages (LRLs) to communicate with each other on the Internet or on some special occasion. In addition, the bilingual professional translator might not good at for using LRLs in comparison to HRLs. The well-known and commonly used off-line translator take advantage of compressing the neural machine translation (NMT) model which trained on huge amounts of data. Therefore, NMT models suffer from the data scarcity problem. Some useful ideas about NMT for LRLs, such as transfer learning [28], word-stem substitution and double-transfer [22] and zero-resource learning [15], have been introduced but many problems still exist in MT. Transfer learning is one of the efficient methods for low-resource NMT task. If we exploit those methods in IAEs with LRLs explicitly, it might not be achieved the more efficient result as HLRs. Generally, the bilingual cognition and some artificial applications (smart devices and machine translation system) are as the integrated union in IEAs. Previous methods almost separated them from each other, and unable to make full use of mutual effectiveness in IEAs.

In this work, therefore, by the comparison to aforementioned approaches, we aim to deal with the problem that how to make full use of these factors and combine them to better develop the international communications in various activities such as international trade, education, and conference etc. Our major idea focuses on investigating some features, mechanisms, and applications of bilingual cognition in IAEs. Exactly, the proposed method is: First, clarify the mother tongue of each international activity, and make some prerequisites before communicating others legitimately. Then, make full use of information and intelligent network tools to bring out the subjective initiative of both parties, while conduct the corresponding research on daily terms and professional terms. Finally, exploit the MT and translation memory tools to better help them make the necessary preparations or exercises. Besides, we have proposed a novel and efficient approach mixed transfer learning (MTL) for low-resource NMT. Our method

achieves outstanding results that when the two parties communicate with each other, they will have some tactic agreement, even they are online or off-line. The HCI better reveal the characteristics, mechanisms, application of bilingual cognition by processing. Experiments on NMT for LRLs, from Arabic (Ar), Farsi (Fa) and Urdu (Ur) to Chinese (Ch) shows that the proposed MTL method is also achieved better results. The key is to take advantage of scientific principles of bilingual informatization and intelligence. Our contributions are as follows:

1. Mitigate the gap between a non-native speaker and native speaker by exploiting prepared or necessary draft.
2. Make full use of the combination of bilingual cognition and artificial translation system in IEAs.
3. Provide some Efficient and effective channels for IEAs.
4. The proposed NMT training approach for LRLs in IEAs is transparent to neural network architecture.

## 2 Background

### 2.1 International Exchange Activities

Intuitively, during the international communication, the neural cognition computing [12, 11] should be one of the necessary parts. International communication is a major activity in an international companys marketing mix. Once a product or service is developed to meet consumer demands and is suitably priced and distributed, the expected consumers must be notified of the products availability and value. International communication [14] consists of those movements which are practiced by the marketer to inform and convince the consumer to purchase. A well-designed advancement mix includes promoting, sales advertisements, particular selling, and public relationships which are mutually augmenting and concentrated on a regular objective.

### 2.2 Neural Machine Translation

Additionally, we also take advantage of machine translation (MT) especially neural machine translation (NMT) in human-computer interaction efficiently and effectively. We take  $X$  as a source language sentence and  $Y$  as a target language sentence, respectively. Given a source sentence  $\mathbf{x} = x_1, \dots, x_i, \dots, x_I$  and a target sentence  $\mathbf{y} = y_1, \dots, y_j, \dots, y_J$ , standard NMT models [25, 1, 27] usually factorize the sentence-level translation probability as a product of word-level probabilities:

$$P(\mathbf{y}|\mathbf{x}; \boldsymbol{\theta}) = \prod_{j=1}^J P(y_j|\mathbf{x}, \mathbf{y}_{<j}; \boldsymbol{\theta}), \quad (1)$$

where  $\boldsymbol{\theta}$  is a set of model parameters,  $\mathbf{y}_{<j}$  is a partial translation. NMT models usually rely on an encoder-decoder scenario.

Let  $\langle \mathbf{X}, \mathbf{Y} \rangle = \{\langle \mathbf{x}^{(n)}, \mathbf{y}^{(n)} \rangle\}_{n=1}^N$  be a training corpus. The log-likelihood of the training parallel data is maximized by the standard training objective :

$$\hat{\boldsymbol{\theta}} = \operatorname{argmax}_{\boldsymbol{\theta}} \left\{ \sum_{n=1}^N \log P(\mathbf{y}^{(n)} | \mathbf{x}^{(n)}; \boldsymbol{\theta}) \right\}. \quad (2)$$

The translation decision rule for unseen source sentence  $\mathbf{x}$  given learned model parameters  $\hat{\boldsymbol{\theta}}$  is given by

$$\hat{\mathbf{y}} = \operatorname{argmax}_{\mathbf{y}} \left\{ P(\mathbf{y} | \mathbf{x}; \hat{\boldsymbol{\theta}}) \right\}. \quad (3)$$

Meanwhile, calculating the highest probability  $\hat{\mathbf{y}} = \hat{y}_1, \dots, \hat{y}_j, \dots, \hat{y}_J$  of the target sentence can be separated at the word level:

$$\hat{y}_j = \operatorname{argmax}_y \left\{ P(y | \mathbf{x}, \hat{\mathbf{y}}_{<j}; \hat{\boldsymbol{\theta}}) \right\}. \quad (4)$$

**Table 1.** The efficient ways for international exchange activities.

Different channels	Example
Social Media Networking	Skype, Facebook, Google+
BLOGS	Twitter, Tumblr, Blogger
Activities Sharing and Storage	Google photos, Flickr, photobucket, instagram
Telecommunication Options	Text follows FaceTime, WhatsApp, Viber, TextPlus, MagicJack

### 3 Methodology

#### 3.1 HCI Method for International Communication

We are staying in an exciting moment that artificial intelligence is becoming ubiquitous and is playing increasingly significant roles in our lives and in the basic infrastructures of science, business, and both social communication and IEAs. We regard that, as shown in Table 1 there were the majority of ways which can be touchable for international communications. However, they just merely play a role of the instrument. We investigate some features and mechanisms of bilingual cognition. As illustrated in Fig. 1, in the first step, confirm and specify what they want to say in IEAs. Namely, the speaker selects the word, and use the phrase to organize the phrase table, then make a sentence by exploiting the phrases. In the confirming the bilingual content, select the bilingual pairs. Exactly, select the semanteme via relatively analogous bilingual semanteme. Finally, revise and update the series of comparison tables. Intuitively, taking advantages of MT to help the speakers and improve the quality of communications.

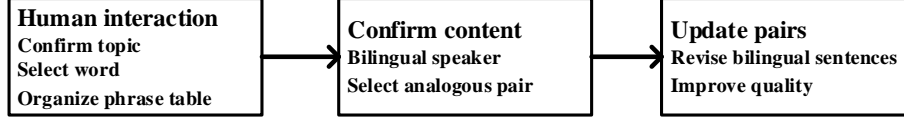


Fig. 1. The architecture of human-computer interaction with bilingual cognition.

### 3.2 Mixed Transfer Learning approach for NMT

For the intention of making full use of the role of the machine and encourage machine help human to improve the quality of their communication efficiency in IEAs. We presented and analyzed the effect of machine translation for international communication as a human-interface toolkit. We follow the major idea of transfer learning (TL) in NMT for LRLs and referred them to incorporate into the architecture of human communication interface system.

We take  $L_3 \rightarrow L_2$  as a parent language pairs and  $L_1 \rightarrow L_2$  as a child language pair.  $L_3$  and  $L_1$  are source languages of parent and child, respectively,  $L_2$  is the target language for both. Additionally, we set the dataset of parent language  $L_3 \rightarrow L_2$  is  $D_{L_3, L_2}$ , while dataset of child language  $L_1 \rightarrow L_2$  is  $D_{L_1, L_2}$ . Besides, we set  $M_{L_3 \rightarrow L_2}$  as the parent language model which learned on parent language dataset  $D_{L_3, L_2}$ . Generally, we initialize the child model  $M_{L_1 \rightarrow L_2}$  by using of parent model [28], and the corresponding parameter of parent model:

$$\theta_{L_3, L_2} = \{ \langle e_{L_3}, W, e_{L_2} \rangle \}, \quad (5)$$

while  $e_{L_3}$  and  $e_{L_2}$  are both source and target embedding of parent model, and  $W$  is their parameters. We also encourage the training objective to maximize the likelihood of dataset  $D_{L_3, L_2}$ :

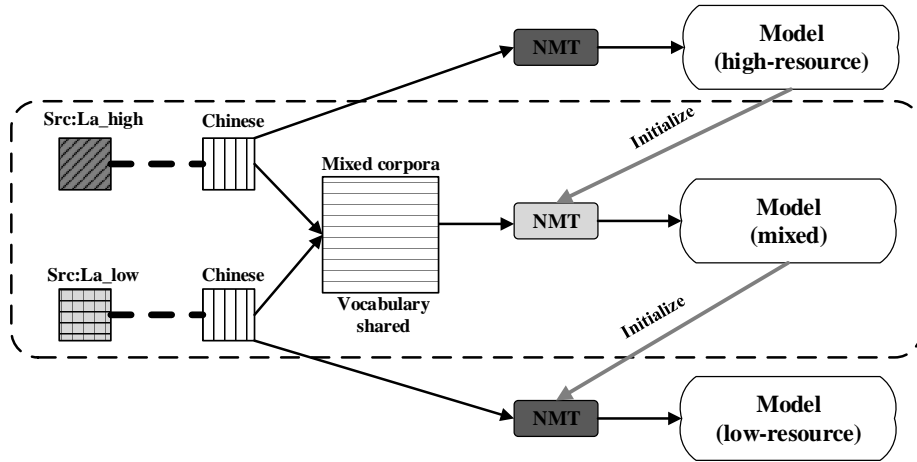
$$\hat{\theta}_{L_3 \rightarrow L_2} = \operatorname{argmax}_{\theta_{L_3 \rightarrow L_2}} \left\{ L(D_{L_3, L_2}, \theta_{L_3, L_2}) \right\}. \quad (6)$$

After that, the child model  $M_{L_1 \rightarrow L_2}$  will be fine-tuned by parent model  $M_{L_3 \rightarrow L_2}$  take advantage of their dataset ( $D_{L_1, L_2}$ ) of child model:

$$\theta_{L_1 \rightarrow L_2} = f(\hat{\theta}_{L_3 \rightarrow L_2}), \quad (7)$$

the learned parameters  $\hat{\theta}_{L_3 \rightarrow L_2}$  of parent model are transferred to child model  $M_{L_1 \rightarrow L_2}$  by initialization function  $f$ .

Intuitively, as depicted in Fig. 2, we inspired by the original transfer learning [28] in NMT and introduce the mixed transfer learning (MTL) approach, which shares the vocabularies between parent and child language. Meanwhile as described in [15], we exploit the oversampling such that data from all language pairs to be of the same size as that of largest language pair, as well as ensuring an equal amount of data per language pair. Then train the mixed model by combining of parent and intermediate model which has been trained on the oversampled mixed bilingual corpus, after that fine-tune the



**Fig. 2.** The architecture of mixed transfer learning for NMT, while the dotted rectangle stands for the training of the mixed model.

**Table 2.** Language features of all languages used in our experiment.

Language	Family	Group	Branch	Order	Unit	Inflection
Arabic (Ar)	Hamito-Semitic	Semitic	South	VSO	Word	High
Farsi (Fa)	Indo-European	Indic	West	SOV	Word	Moderate
Urdu (Ur)		Iranian	Iranian	SOV	Word	Moderate
Chinese (Ch)	Sino-Tibetan	Chinese	Sinitic	SVO	Character	Light

**Table 3.** Characteristics of parallel corpora. While “Vocab.” and “# Word” represent vocabulary (word type) and word token, respectively. Besides, “#Cov.” stands for corresponding covering rates of each language.

Languages	Train	Dev	Test	Source			Target		
				Vocab.	# Word	#Cov.[%]	Vocab.	# Word	#Cov.[%]
Ar → Ch	5.1M	2.0K	2.0K	1.0M	32.2M	88.30	0.5M	37.4M	96.80
Fa → Ch	1.4M	2.0K	1.0K	0.2M	10.4M	96.80	0.2M	10.0M	96.80
Ur → Ch	78.0K	2.0K	1.0K	17.6K	2.6M	100.00	12.7K	2.4M	100.00

child model to improve their translation quality, and mitigate the cost of professional translators, and better help to international communication during the using of LRLs with each other.

## 4 Experiments

### 4.1 Setup

All the datasets which are used in this work publicly available on Open Subtitle2016<sup>1</sup> and Tanzil corpora<sup>2</sup>. The corpus feature and specifications are shown in Table 2 and Table 3, respectively, in this corpora the target side is identically set with Chinese, while source side is different parent RRLs or child LRLs. Additionally, parent language pairs Ar  $\rightarrow$  Ch, Fa  $\rightarrow$  Ch are collected from Open Subtitle2016 corpora, and the child language pair Ur  $\rightarrow$  Ch is obtained from Tanzil corpora.

In the preprocessing step, we used NIUTTRANS preprocessing *perl script*<sup>3</sup> to remove and clean illegal Chinese parallel sentences (our target side) from original corpus. Besides, we also prepared several preprocessing *python scripts* for both source and target side. Each of these scripts mainly works for LRL and Chinese parallel corpus, as well as re-cleaning after removing illegal characters, removing the blank lines, removing illegal symbols and double checking non-Chinese characters, and converter which is used for converting simplified Chinese and traditional Chinese. We also use open source Chinese word stemmer system THULAC<sup>4</sup> for Chinese language [17]. Meanwhile, we exploited *tokenizer.perl* toolkit<sup>5</sup> which was provided by state-of-the-art (SOTA) phrase-based statistical machine translation (SMT) system MOSES [16] for word tokenization. Moreover, we report results without any UNK-replacement techniques [19].

Additionally, in our full experiment, we make use of the attention-based encoder-decoder model which gated recurrent unit for NMT system DL4MT<sup>6</sup>. Likewise, we run the experiments approximately for 3–4 days except for finetuning (only 2–3 days) on single NVIDIA TITAN X (PASCAL) GPU almost with default parameters of DL4MT, just slightly modified some of the dimensions, such as word embedding is 620, hidden state embedding is 1000, we limit the vocabularies to the most frequent 30K words (covering rates as shown in Table 3) and batch size is 80, sentence max length is 50. Basically, in our experiments the baseline and our method in terms of case-insensitive BLEU<sup>7</sup> scores [21].

### 4.2 Effect of MTL for low-resource NMT

Intuitively, it is tractable to infer from Table 2 that the parent languages Ar and Fa both of them are similar with child language Ur, likewise the Fa and Ur belong to same language family. We referred to the relatedness and explore the shared word rate among

<sup>1</sup><http://opus.nlpl.eu/OpenSubtitles2016.php>

<sup>2</sup><http://opus.nlpl.eu/Tanzil.php>

<sup>3</sup><http://www.nlplab.com/niuplan/NiuTrans.YourData.html>

<sup>4</sup><https://github.com/thunlp/THULAC-Python>

<sup>5</sup><https://github.com/moses-smt/mosesdecoder/tree/master/scripts/tokenizer/tokenizer.perl>

<sup>6</sup><https://github.com/nyu-dl/dl4mt-tutorial>

<sup>7</sup><https://github.com/moses-smt/mosesdecoder/blob/master/scripts/generic/multi-bleu.perl>



**Table 4.** The shared word rate between Ar, Fa, and Ur, and they are converted into Latin scripts.

	Arabic	Farsi	Urdu
Arabic		12.36%	8.61%
Farsi	2.51%		8.62%
Urdu	0.16%	0.78%	

**Table 5.** Effect of shared vocabulary single fine-tuning. While parent language pairs with the token “(Shared)” represents they shared same vocabularies with child language pair. Conversely, language pairs with the token “(Non-Shared)” stands for using own vocabularies rather than shared vocabulary anymore. “++”: significantly better than RNNSEARCH ( $p < 0.01$ ).

Method	Parent	Child	BLEU $\uparrow$
RNNSEARCH	N/A	Ur $\rightarrow$ Ch	18.31
MIXED	Fa $\rightarrow$ Ch (Non-Shared)	Ur $\rightarrow$ Ch	19.73
	Ar $\rightarrow$ Ch (Non-Shared)		20.04
	Fa $\rightarrow$ Ch (Shared)		<b>21.69<sup>++</sup></b>
	Ar $\rightarrow$ Ch (Shared)		<b>22.44<sup>++</sup></b>

them (see Table 4). In this experiment, we take RNNSEARCH as a baseline which is standard attention based encoder-decoder model [1] for NMT.

Generally, inspired by [18] using shared vocabulary between general domain corpus and an in-domain corpus to improve translation quality of the in-domain model. In this work, we have also shared vocabulary between parent language pairs (Ar  $\rightarrow$  Ch and Fa  $\rightarrow$  Ch) and child language pair (Ur  $\rightarrow$  Ch). First, pre-train the low-resource NMT, then combine both source sides (Ar/Fa, Ur) and target sides (both of two groups are Ch) of parent and child sentences to create the big corpus, as well as mixed parent and child corpora. Then generate shared vocabulary, it includes both higher frequency words of parent and child language pairs, to train the mixed model. Finally, Initialize the Ur  $\rightarrow$  Ch via Ar  $\rightarrow$  Ch and Fa  $\rightarrow$  Ch with and without shared vocabularies, respectively. Exactly, train the parent model  $M_{Ar \rightarrow Ch}$  and  $M_{Fa \rightarrow Ch}$  with their own vocabularies which only consisted of Ar or Fa, and with mixed vocabularies that include both of parent and child words to train corresponding parent models, then initialize the same child model  $M_{Ur \rightarrow Ch}$ . As given in Table 5, proposed approach obtained improvements exploiting Ar  $\rightarrow$  Ch (shared) and leveraging Fa  $\rightarrow$  Ch (shared) comparison with non-shared (used own vocabulary) vocabulary parent model.

## 5 Related Work

As we have aforementioned, there were many methods for developing international communication [6], international education [23], international trade [2], intercultural communications [8], and intercultural business communication [4]. Moreover, the variety of efficient ideas and approaches have been introduced, however, many factors and essential parts were neglected. Intuitively, we can make full use of some features and architectures by leveraging artificial intelligent mechanisms. Likewise, we regard

that the consideration of the mother tongue of speakers and combinations of bilingual cognition and machine translation could better help and improve the quality of IEAs.

In the literature, the transfer learning (TL) [26] method has been being widely used in computer vision [20] and domain adaptation [3, 5]. We have also referred the TL and incorporate them into NMT [7, 10, 9, 24] to better help the training procedure of LRLs by leveraging of highly analogous HRLs. Additionally, many researchers [28, 22] have also paid attention to LRLs NMT, however, neglected the sharing vocabularies between similar or highly related HRLs and LRLs. In the IEAs, the majority of users explicitly exploit the original translation system which was trained on the big parallel corpus. In this case, the developing procedure becomes more sophisticated that preparing huge amounts of data and training them for several hours on GPUs or even TPUs. While some others also try to improve the quality of translation via leveraging the word substitution [22] between same language family, group even the same language branch. But it has some limitations between other similar languages which are in various language family or group. Likewise, Google [15] also introduce some useful methods for zero-resource translation by combining HRLs and LRLs together with oversampling. However, their approach ignores the relatedness of languages and make some turbulence signals when LRLs receiving parameters from HRLs during training.

## 6 Conclusion and Future Work

In this work, we discuss the efficient architecture of HCI with bilingual cognition. For the intention of, how to make full use of some factors and combine them to better develop the HCI, our method mainly focuses on the investigation of IEAs by using artificial intelligent mechanisms. Besides, we have also proposed a novel training method mixed transfer learning (MTL) for NMT, which is used in IEAs as the speaker needs to use LRLs. Additionally, instead of exploiting the original NMT model, we leverage the MTL method to share vocabulary to train the NMT model. Then guide the LRLs model by highly related HRLs model, and mitigate computation space and reduce memory consumption and time. In the future work, we plan to further validate the effectiveness of our method on more NLP and other tasks except for IEAs, meanwhile try to leverage on morphologically poor languages.

## References

1. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. In: Proceedings of ICLR (2015)
2. Bidabad, B.: A convention for international trade (based on islamic sufi teachings). International Journal of Law and Management (2015)
3. Carpuat, M., Daumé III, H., Henry, H., Irvine, A., Jagarlamudi, J., Rudinger, R.: Sensespotting: Never let your parallel data tie you to an old domain. In: Proceedings of ACL (2013)
4. Chien, T.C.: Intercultural training for taiwanese business expatriates. Industrial and Commercial Training (2012)
5. Chu, C., Dabre, R., Kurohashi, S.: An empirical comparison of simple domain adaptation methods for neural machine translation. CoRR [abs/1701.03214](#) (2017)

6. Cristian Chelariu, T.O.: "communication technology in international business-to-business relationships. *Journal of Business & Industrial Marketing* (2014)
7. Dong, D., Wu, H., He, W., Yu, D., Wang, H.: Multi-task learning for multiple language translation. In: *Proceedings of ACL* (2015)
8. Eila Isotalus, M.L.K.K.: Ethics and intercultural communication in diversity management. *Equality, Diversity and Inclusion: An International Journal* (2018)
9. Firat, O., Cho, K., Bengio, Y.: Multi-way, multilingual neural machine translation with a shared attention mechanism. In: *Proceedings of NAACL* (2016)
10. Firat, O., Sankaran, B., Al-Onaizan, Y., Vural, F.T.Y., Cho, K.: Zero-resource translation with multi-lingual neural machine translation. In: *Proceedings of EMNLP* (2016)
11. Gerard A.J. Pounder, Ruel L.A. Ellis, G.F.L.: Cognitive function synthesis: preliminary results. *Kybernetes* (2017)
12. Glenn Finch, Brian Goehring, A.M.: The enticing promise of cognitive computing: high-value functional efficiencies and innovative enterprise capabilities. *Strategy & Leadership* (2017)
13. Grosjean, F., Gremaud-Brandhorst, J., Grosjean, L.: The bilingual s language modes (1999)
14. Johnson, G.J.: The international encyclopedia of digital communication and society. *Reference Reviews* (2017)
15. Johnson, M., Schuster, M., Le, Q.V., Krikun, M., Wu, Y., Chen, Z., Thorat, N., Viégas, F., Wattenberg, M., Corrado, G., Hughes, M., Dean, J.: Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association of Computational Linguistics* (2017)
16. Koehn, P., Och, F.J., Marcu, D.: Statistical phrase-based translation. In: *Proceedings of NAACL* (2003)
17. Li, Z., Sun, M.: Punctuation as implicit annotations for chinese word segmentation. *Computational Linguistics* **35**, 505–512 (2009)
18. Luong, M.T., Manning, C.D.: Stanford neural machine translation systems for spoken language domains. In: *Proceedings of IWSLT* (2015)
19. Luong, T., Sutskever, I., Le, Q.V., Vinyals, O., Zaremba, W.: Addressing the rare word problem in neural machine translation. In: *ACL* (2015)
20. Oquab, M., Bottou, L., Laptev, I., Sivic, J.: Learning and transferring mid-level image representations using convolutional neural networks. *2014 IEEE Conference on Computer Vision and Pattern Recognition* pp. 1717–1724 (2014)
21. Papineni, K., Roukos, S., Ward, T., Zhu, W.J.: Bleu: a methof for automatic evaluation of machine translation. In: *Proceedings of ACL* (2002)
22. Passban, P., Liu, Q., Way, A.: Translating low-resource languages by vocabulary adaptation from close counterparts. *ACM Trans. Asian & Low-Resource Lang. Inf. Process.* **16**, 29:1–29:14 (2017)
23. Petrina M. Davidson, Calley Stevens Taylor, M.P.N.D.A.W.W.: Introduction reflecting on trends in comparative and international education: A three-year examination of research publications. *International Perspectives on Education and Society* (2018)
24. Sennrich, R., Haddow, B., Birch, A.: Neural machine translation of rare words with subword units. *CoRR* **abs/1508.07909** (2016)
25. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. In: *Proceedings of NIPS* (2014)
26. Tan, B., Zhang, Y., Pan, S.J., Yang, Q.: Distant domain transfer learning. In: *AAAI* (2017)
27. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: *Proceedings of NIPS* (2017)
28. Zoph, B., Yuret, D., May, J., Knight, K.: Transfer learning for low-resource neural machine translation. In: *Proceedings of EMNLP* (2016)