Keeping the Meanings of the Source Text: An introduction to Yes Translate

Xiaoheng Zhang

Department of Chinese and Bilingual Studies, Hong Kong Polytechnic University
ctxzhang@polyu.edu.hk

Abstract. The primary task of language translation is to faithfully pass the meaning(s) of the source text to the target language. Unfortunately, meanings often get lost or distorted in machine translation, including state-of-the-art Google Translate and Baidu Translate. Yes Translate is a Chinese-English translation tool to be maximally loyal to the source text while maintaining adequate fluency. This is implementable by avoiding risky actions of word deleting, adding and re-ordering. The tool is supported by an 116,000-words Dictionary. In an experiment on natural news articles freely selected by themselves, 10 postgraduate students with good command of Chinese and English all agreed or strongly agreed that the general meaning of the translation by Yes Translate was correct and understandable. And 9 out of the 10 students agreed or strongly agreed that the general meaning of each sentence was correct.

Keywords: Machine translation, meanings, Yes Translate, YES-CEDICT

1   Introduction

The basic goal of translation is to faithfully convey the intended meaning(s) of the author in one language to the reader in a different language. However, there is large room for improvement in this aspect in machine transition (MT) (Bhattacharyya, 2015, p30), which is true even to the most popular and state-of-the-art systems such as Google Translate and Baidu Translate, as demonstrated by the following example.

The source text is a Chinese article from Yahoo News (Yahoo, 2016). Due to limitation of space, only the title and one paragraph is presented here, as shown in Figure 1.

教宗訪希臘難民營　帶走12敘利亞難民返梵蒂岡
星島日報 – 2016年4月17日星期日上午11:33.

…
　　教宗是於一星期前決定應梵蒂岡一名官員建議，收容3個敘利亞家庭。教宗解釋，他接納建議，是因為符合他到訪萊斯沃斯島的精神。他說：「這好比滴了一個水滴到大海，但這個小水滴已令到大海變得不一樣。」
…

Fig. 1. Source text in Chinese

English translations by Google and Baidu are shown in Figure 2 and 3.

Pope to visit Greece 12 Syrian refugees camps took back the Vatican
Sing Tao Daily - April 17, 2016 11:33 Sunday morning.
...
Vatican Pope decided should an official recommendation, accommodating three families in Syria a week ago. Pope explained that he accepted the proposal, because in line with the spirit of his visit Lesbos. He said: "It's like a drop of water onto the sea, but the sea have caused the water droplets become different."
...

Fig. 2. Translation by Google Translate (https://translate.google.com.hk/, May 28, 2016)

The Pope's visit to the Greek camp took 12 refugees from Syria back to the Vatican
Sing Tao Daily - April 17, 2016 Sunday morning 11:33.
...
The Pope is in a week before the decision should be a Vatican official suggested that the 3 Syria family housing. The Pope explained that he accepted the proposal because he visited with the spirit of lesbos. He said: "this is like a drop of water to the sea, but the small water droplets have made the sea become different. "
...

Fig. 3. Translation by Baidu Translate (http://translate.baidu.com/, May 28, 2016)

There are several problems in the translations.

(1) Loss of words and their meanings. For example, in Baidu's translation of the second sentence in the paragraph, Chinese word "符合" (in line with) in the source text is simply ignored. Another example is an error made by both Google and Baidu. The meaning of one of the two occurrences of word "難民" (refugee) in the original title is missing in the translation, while the function of the space in the Chinese title is also neglected, bringing more confusion to the reader.

(2) Loss of the intended meaning of a word. For example, in the first sentence of Google's translation, when clicking word "should" on Google Translate, we can see that it corresponds to word "應" in the source text. That is obviously wrong translation. And, to make it worse, the correct translation (such as "accept") is not available on the candidate list. Baidu makes similar mistakes.

(3) Adding words with non-intended meanings. For example, in the title of the article, Google Translate added a "to" before "visit" to form "Pope to visit Greece…", meaning the visit was going to happen, while actually it had

already happened. In addition, Google and Baidu both incorrectly use plural "droplets", while the source text indicates a single drop (by both 一個水滴 and 這個小水滴).

(4) Improper words order. For example, in the last sentence of Google's translation, improper movement of words has generated "but the sea have caused the water droplets become different", contrary to the intended meaning of "but the water droplet has caused the sea become different." of the source text. And in Google's translation of the title, "12 Syrian refugees" has been moved to a wrong location. The correct place is after "took". Google also incorrectly moved "Vatican" away from its modification of "an official" to modify the pope in the first sentence.

A post-edited version of the translation by Google is presented in Figure 4.

Pope ~~to~~ visit<u>ed</u> Greece ~~12 Syrian~~ refugee~~s~~ camps <u>and</u> took <u>12 Syrian refugees</u>
back <u>to</u> the Vatican
Sing Tao Daily - April 17, 2016 11:33 Sunday morning.

...

~~Vatican~~ Pope decided ~~should~~ <u>to accept a</u>~~n~~ <u>Vatican</u> official<u>'s</u> recommendation~~,~~ <u>of</u> accommodating three families ~~in~~ <u>from</u> Syria a week ago. Pope explained that he accepted the proposal, because<u> it was</u> in line with the spirit of his visit <u>to</u> Lesbos. He said: "It's like a drop of water onto the sea, but the <u>water droplet</u> ~~sea~~ ha<u>s</u>~~ve~~ caused the <u>sea</u> ~~water droplets~~ become different."

...

Fig. 4. A post-edited version of the translation by Google

In the following sections, we will introduce Yes Translate, a Chinese-English translation tool aiming to be maximally loyal to the meanings of the source text while maintaining adequate fluency, and be free from the above-mentioned problems.

2   System Design Considerations

The purpose is to build a Chinese-English bidirectional machine translation (MT) tool for general purposes. We selected English and Chinese because we are more familiar with these two languages and use them more frequently. According to the Internet World Stats (2016), English and Chinese are by far the two most popular languages on the Web. Their Internet users are 872,950,266 and 704,484,396 respectively, followed by a large gap before the 3th-place Spanish with 256,787,878 users. We are working on Chinese-to-English translation first because a large Chinese-English dictionary of over 110,000 words and phrases is available (Zhang, Li and Lun, 2015).

Keeping source text meanings in machine translation is implementable by avoiding risky actions of word deleting, adding and re-ordering, and by employing more reliable methods.

Every word in the source text will be taken into consideration, not ignored without good reasons. For the example of the previous section, the meanings of word "符合" (in line with) and all uses of "難民" (refugee) will be kept in the translation, while the function of the space in the Chinese title is kept as a mark between two phrases.

All possible meanings and functions of a word or phrase will be maintained. For the example of word "應", in addition to "should", "accept" is added to the translation candidate list. Possible translations of a word will be sorted according to frequency of use. For example, in CC-CEDICT, the first sense/definition of 届 is "arrive". However, on corpus 现代汉语语料库 (http://www.cncorpus.org/index.aspx), the first 20 entries of the concordance are all in the form of "第 (number) 届". Hence we should put its corresponding English translation to the top. We need a frequency list of each meaning (or translation) of each word (translation unit). And if the context is also considered, the correctness rate may be further improved. For example, word "将" has two frequently used functions of "representing a future tense" and "indicating the following is the object of an action". From the Chinese corpus, we can see that if "将" is followed by a verb then "will" is a more possible translation, and object indicating is more likely when followed by a noun phrase. For example, "将吃完饭" and "将饭吃完" will be translated into "will finish eating food" and "(object:) food finish eating" respectively, while Google Translate output "The meal"(将吃完饭) and "will eat rice." (将饭吃完). That means optimization can be achieved using big data and other resources.

Addition of words or letters with non-intended meaning should be avoided. For the news article in Section 1, it is better not to add a "to" before "visit" in the title of the article, or add an "s" at the end of "droplet" in the last sentence.

Movement of words is not allowed when it may incorrectly change the meaning of the sentence. Then for the pope's news, we will have the correct translation of "but the water drop has caused the sea become different.", "took 12 Syrian refugees back to the Vatican" and "a Vatican official's recommendation". Moving a word or not may require different expressions. For example, Chinese word "之后" is often translated to "after" in English. Phrase "3年之后" is translated into "after 3 years". It involves moving word 之后 (after) to the front of 3年 (3 years), which may be risky for the computer if 3年 is replaced by a more complicated phrase. If the word is kept in its original position, then "3年之后我去英国 (after 3 years I went to UK)" will be translated into "3 years after I went to UK", which is more likely to mean "3 years after my going to UK". A safer option is to translate 之后 into "later" or "afterwards", or add a comma after "after", then "3年之后我去英国" will be translated into "3 years later I went to UK" , "3 years afterwards I went to UK" or "3 years after, I went to UK", all can convey the correct meaning without words movement.

The output text may be more similar to the source text in word order, however that does not prevent passing the correct meanings to the target language, as shown by the feasibility of simultaneous interpretation and sight translation.

Translation (in both direction) is carried out sentence by sentence in order of the original text. Starting from the beginning of a sentence, the computer will repeat two processing stages:

(1) Select the next translation element in the source text. Here we employ the maximum-length-first word segmentation method. This is implemented by sorting the bilingual dictionary by word (and phrase) lengths in descending order, before matching with the source text.

(2) Select a translation of the element in the target language. If there are more than one meaning (or translation) of the element, the most possible (frequently used and reliable) one is presented, while the rest are preserved in a backup candidate list for the user to select once the translation presented is not consistent with the context.

There are three levels of goals:
1) Keeping all the possible meanings (most important)
2) Using correct grammar
3) Achieving native speaker fluency

We would first of all focus on successful convey of meaning, because that is the basic requirement of the reader. All the possible meanings of the author should be well kept in the translation output, for the reader's reference. The feasibility of the design is high, because if the source text is understandable to its readers, so will a translation honestly keeping all its meanings (be understandable) to its readers in the target language.

3　Implementation and Improvement

3.1　Implementation

Yes Translate is implemented in the form of a web site for convenience of accessibility. The initial state is shown in Figure 5.

Fig. 5. The initial state of Yes Translate

To translate a text from Chinese to English or vice visa, simply type or paste it in the source text window on the left, then click a translating button at the top (English-to-Simplified Chinese, English-to-Traditional Chinese, Simplified Chinese to English, and Traditional Chinese to English), and the translation will appear in the target text widow on the right hand side. And in addition to the PC version, a mobile phone version is available, with the source and target text windows arranged in a vertical mode.

Internally, Yes Translate has four components: the Chinese-English dictionary (or called translation memory), the English-Chinese dictionary (translation memory), the translation engine and the user interface.

Both the Chinese-English dictionary and the English-Chinese dictionary are developed from the latest version of YES-CEDICT (Zhang and Li, 2016), which includes over 116,000 word entries, covering the following (except a small number of characters not displayable on the computer):

- All the 5,000 words of the new word list of HSK (Hanyu Shuipin Kaoshi), the official Chinese Proficiency Test of China mainland (Hanban, 2012);
- All the 8,000 words of the new word list of TOCFL (Test of Chinese as a Foreign Language), the official Chinese proficiency test of Taiwan (National Taiwan Normal University, 2013);
- All 113,935 headwords in CC-CEDICT, the most popular downloadable Chinese dictionary on the Web (MDBG, 2016);
- All 11,408 characters in the Complete List of Chinese Characters Used by the Media in 2013 (2013 年度媒体用字总表, National Language Commission of China, 2014);
- All 13,000 plus characters in the latest version of the Xinhua Dictionary (Linguistic Institute of the Chinese Academy of Social Sciences, 2011);
- All 13,000 plus characters in the latest version of the Contemporary Chinese Dictionary (Linguistic Institute of the Chinese Academy of Social Sciences, 2012).

Other references include the Oxford Chinese Dictionary (Kleeman and Yu, 2010), ABC Chinese-English Comprehensive Dictionary (DeFrancis, 2003), the List of Frequently-used Words of Modern Chinese (in frequency order, National Language Commission of China, 2008), big data (such as corpora on the Web), Google Translate, Baidu Translate as well as our working experiences.

An entry in the English-Chinese dictionary (translation memory) has the contents of English word (or phrase), translation in Simplified Chinese, translation in Traditional Chinese and linguistic labels, and the entries are sorted by the length of the English words descendingly. An entry in the Chinese-English dictionary (translation memory) has the contents of Simplified Chinese, Traditional Chinese, translation in English and linguistic labels, and the entries are sorted by the length of the Chinese words descendingly. Linguistic labels include marks of part of speech, time, voice, plurality, etc,

If a word has multiple meanings (or translations), the meanings are sorted according to their frequency of use. For example, in CC-CEDICT, Chinese word 不满 is defined as: /resentful/discontented/dissatisfied/. But in Google Translate, the frequency order is /dissatisfied/discontented/resentful/. Hence, the frequency order is adopted.

Some translations are added to the existing words in the dictionary for MT purpose. For example, Chinese word 只有 is often translated to "only" referring to the only condition for something to happen, such as "我们只有努力才能成功" (We will succeed only by working hard). We have added the possible phrasal meaning of "only have", such as in "我们只有一个地球。" (We only have one earth.)

Capitalization of the first letter of a sentence in the English target text is fulfilled in a simple way: the initial letter of a word is capitalized if it is preceded by a next line (/n) or a full stop (.) character.

No repeating English translation will appear in the option list. For example, to look after the variant words of 裏 and 裡 in traditional Chinese-to-English translation, there are

inside	里	裏
inside	里	裡
lining	里	裏
lining	里	裡

among the entries of simplified Chinese 里 in the Chinese-English dictionary. In simplified Chinese to English translation. only

| inside | 里 |
| lining | 里 |

will appear in the optional list.

Due to limitation of space here, more technical details will be reported in a separate paper.

3.2 Improvement

Substantial improvement of the software has been achieved by continuous tests on real-life Chinese texts. In the recent months, we have used Yes Translate to translate the complete government work report by Premier Li Keqing (2016), some paragraphs from the Bible and a large number of news articles from China Mainland, Taiwan, Hong Kong and other places on Google, Baidu, BBC, VOA, etc. Weak points of translation revealed by the tests are carefully analyzed and effectively solved.

For example, Yes Translate used to translate Chinese word 把 into English "handle" or "hold", while in real life, it is most frequently used as a preposition in the "把+O+V" expressions (with particle 把 marking the beginning of an object noun phrase followed by the verb phrase), which are very common in Chinese but not used in English. Translating an "S+把+O+V" sentence to "S+V+O" sentence is likely to lose emphasis on the object. A more serious risk is caused by incorrect detecting of the boundaries of the following noun phrase or the verb phrase, which is more likely to happen when the object is long. For instance, Google Translate translates:

"把去北京的人帶回香港。" into a mess of

"The man went back to Hong Kong to Beijing."

To solve this problem, we have added a new entry to the dictionary to translate 把 into an "O+V" marker, avoiding movement of the object NP. A much better translation is generated for the pervious sentence:

"(object+vt:) Go (to) Beijing ('s) people bring back to Hong Kong."

By human intelligence, the reader should have no difficulty finding out the O-V boundary between "people" and "bring". Word 將 can be processed in a similar way when functioning as 把. Translate to "(object+vt:)" when followed by a noun, and into "will" when followed by a verb, while keeping the other possibilities on the option list.

的 is the most frequently used word in modern Chinese. It can be translated into "of", "that", "'s" etc. in English. For instance, "我们老师的办公室" (the office **of** our teacher), and "我将要读的书" "the book **that** I will read". Both "of" and "that" involve movement of the attributive, which can be risky. Translation with "'s" follows the same word order of Chinese. And "'s" is not limited to express possession, e.g., "a master's degree", "block sb.'s path", "New Year's Day", "a lifetime's absence", "father's elder brother's wife", etc. Hence, it seems safer to use ('s) before a more reliable grammar parser is available. The previous two sentences will then be translated into "our teacher's office", and "I will read's book". The second sentence may not sound grammatically correct, but its meaning is more accurate than Google's translation: "I am going to read the book".

4   Evaluation

Yes Translate is on the Web for testing. The initial results are very encouraging.

4.1   Translation of the pope's news article

Figure 6 is a translation by Yes Translate of the news article on Pope's visit to the refugee camp. For convenience of reading, the Chinese source text of Figure 1 is reprinted here.

教宗訪希臘難民營　帶走12敘利亞難民返梵蒂岡
星島日報 – 2016年4月17日星期日上午11:33.

…
　　教宗是於一星期前決定應梵蒂岡一名官員建議，收容3個敘利亞家庭。教宗解釋，他接納建議，是因為符合他到訪萊斯沃斯島的精神。他說：「這好比滴了一個水滴到大海，但這個小水滴已令到大海變得不一樣。」
…

Fig. 1. Source text in Chinese

```
The pope visit Greece refugee camp    take away 12 Syria refugee return (to)
Vatican City

Sing Tao Daily - 2016 year April 17 date Sunday morning 11:33.

...

The pope be in one week before, decide should Vatican City one name officer
suggest , house 3 Syria family . The pope explain , he accept suggest , be
because satisfy he visit Lesbos Island ('s) spirit . He say : " this be just
like drop (completed) a drop of water to (the) sea , but this small drop of
water already make (the) sea become different . "

...
```

Fig. 6. Translation by Yes Translate
(May 28, 2016, http://www.mypolyuweb.hk/ctxzhang/yesmt/)

Comparing with the translation of Google and Baidu, the meanings of the source text are better preserved by Yes Translate. And assistance to post-editing is available: If a translation element has more than one option, it is presented in the form of an interactive link, which, when clicked, will display the candidate list of translations for the reader to select. For example, the word "should" in the first sentence is obviously inconsistent with the context. When clicked, it will present "should, answer, accepted, etc", among which "accept" is selected. Similarly, verb "suggest" has been changed into a noun form, as shown in Figure 7.

```
The pope visit Greece refugee camp    take away 12 Syria refugee return (to)
Vatican City

Sing Tao Daily - 2016 year April 17 date Sunday morning 11:33.

...

The pope be in one week before, decide accept Vatican City one name officer
suggestion , house 3 Syria family . The pope explain , he accept suggestion
, be because satisfy he visit Lesbos Island ('s) spirit . He say : " this
be just like drop (completed) a drop of water to (the) sea , but this small
drop of water already make (the) sea become different . "

...
```

Fig. 7. Improvement by computer assisted proofing

With minimum editing, the text has become more accurate and intelligible, though a number of grammar errors still exist.

4.2 Evaluation by students

In this experiment, 10 postgraduate students from our class of "Computer Tools for the Language Professionals" were asked to select a simplified or traditional Chinese news article of their own interest from the Web, and translate it into English with Yes Translate. Then personally proofread the translation and fill the questionnaire. The results are summarized in Table 1.

Table 1. Evaluation of Yes Translate's translation by 10 students

	1 strongly disagree	2 disagree	3 neutral	4 agree	5 strongly agree	Average mark
1 General meaning of the text correct and understandable.				7	3	4.3
2 General meaning of each sentence correct.			1	8	1	4
3 Grammar is OK.		1	2	4	3	3.9
4 Writing is OK.			1	3	6	4.5
5 The tool is useful.				3	7	4.7

The accuracy of the translation is quite exciting. Of all the 10 students, 7 agree and 3 strongly agree that the general meaning of the English translation is correct and understandable. The average mark is (7*4+3*5)/10=4.3, of a full mark of 5. And 9 out of the 10 students agree or strongly agree that the general meaning of each sentence is correct. The fluency of the translation is roughly OK, as reflected by the marks of grammar and writing, though the marks may be lower if we changed "OK" into "good" in questions 3 and 4. Finally all students agree or strongly agree that the tool is useful, with an impressive average mark of 4.7.

4.3   Evaluation by the computer

A small-scaled evaluation in BLEU and NIST has been performed on Asiya-Online (http://asiya.lsi.upc.edu/demo/asiya_online.php). Data for the source and reference texts are 5 articles from VOA Chinese-English bilingual news of July 25 to 27, 2016 (on http://www.voachinese.com/z/2404.html). There are totally 1,722 characters in the Chinese source file, and 967 words in the English Reference file. Table 2 presents the BLEU and NIST values of system translations by Google Translate, Baidu Translate and Yes Translate.

Table 2. BLEU and NIST values of translations by Google, Baidu and Yes.

	BLEU	NIST	words
Google	0.1496	4.4727	922
Baidu	0.1462	4.1836	979
Yes	0.0175	1.7821	1312

The BLEU and NIST values of Yes are much lower than Google and Baidu. And all three systems score less than 50% of the full marks of BLEU (1) and NIST (10). Possible reasons include

- The Yes translation text is much longer than the reference text, (1312-967)/967=35.7%.
- Yes uses expressions which are not used in traditional English, for example "('s)", "(completed)", "(object + vt:)", etc.
- Frequently used function words such as "the", "a" and "an" are often missing in Yes.

Yes Translate is still very young. Formal development of the software started less than 7 months ago. There is large room for improvement.

As a matter of fact, there are excellent translations which have been scored unbelievably low BLEU and NIST marks. For instance, the Bible in Basic English on the bible website at http://www.o-bible.com/kjv.html is a high-quality translation. However, according to our experiment on Asiya-Online with the King James Version for reference, the BLEU value given to Chapter 1 of Genesis in the Basic English version is 0.4565, below half the full mark of 1.

5   Conclusion and Further Development

Meanings of the source text often get lost or distorted by machine translation tools, including leader products such as Google Translate and Baidu Translate. Yes Translate is a Chinese-English translation tool to be maximally loyal to the source text while maintaining adequate fluency. In an experiment on 10 students with good command of Chinese and English, all agreed or strongly agreed that the general meaning of the English translation by Yes Translate was correct and understandable. And 9 out of the 10 students agreed or strongly agreed that the general meaning of each sentence was correct.

Other important features of Yes Translate are its effective support to post-editing, and the client-side work mode. That means no continuous connection to the server is required. The size of the whole website is 12.6MB, very handy for a machine translation system.

There is large space for further improvement. Some sentences are still translated without accurate meanings. The mistakes are mostly caused by incorrect word segmentation, for example,

一张/人手/大小的试纸 (One sheet (of paper) **manpower** size ('s) test paper) ,

/在理/论方面 (**Reasonable** theory aspect),

国际债权/方向/希腊发放更多的救助款项 (International creditor's rights (law) **direction** Greece provide more ('s) aid funds).

Ambiguous word segmentations in the form of "cc" vs "c/c" is limited to the two characters, hence can be properly represented by putting both word "cc" and phrase "c/c" in the dictionary. For example, both "人手, manpower" and "人手, human hands" are included.

Ambiguity in the form of "cc/c" vs "c/cc" can be dealt with in a similar way. For example, the error segmentation of /在理/论方面 can be avoided by adding phrase

"在理论, in theory" to the translation dictionary. Unfortunately, there are too many words starting with 理 which may follow 在, including 在理论…, 在理解…, 在理想…, etc. And the ambiguous chain can be longer than two words, e.g.

/在理/大校/园 [/在/理大/校园/].

It seems more cost-effective to delete entry 在理 from the dictionary. Then 在理 will be translated into "in reason", which is also acceptable. However, some words are too popular to delete, e.g., 十分, 有力, in spite of the existence of expressions like 十分钟, 有力不从心的时刻.

A better method is to further improve word segmentation. For example, we can take advantage of the feature that, comparing with "十分 (ten seconds/marks)", "十分 (very)" is more likely to be followed by an adjective or adverb. Another method is to combine forward word segmentation processing and backward processing, and mark the inconsistent parts for post-editing choices. But that will slow down the translation substantially.

However, word segmentation can never be 100% correct (Zhang and Sun, 2012). When probability is involved, there is space of missing the correct choice. A more ideal way is to promote word segmented writing (Chen, 1996; Zhang 1998) like English, but that will take a very long period of time. Yes Translate is already capable to process word-segmented text. For example, "在 理大 校园" is translated to "at PolyU campus".

Another aspect for further improvement is grammar and fluency of the output. Grammar errors are mainly caused by the lack of physical marks in the Chinese language to represent differences in time, voice, number, phrase structures, etc., and due to the limited time we have been working on this project.

We will continue to improve Yes Translate on real-life texts, especially daily news articles on various topics. Yes Translate is mainly data-driven, improvement of the system heavily relies on improvement of the dictionary.

Work is now focused on the Chinese-English part. The next step is to improve English-Chinese machine translation.

References

1. Bhattacharyya, P.: Machine Translation. Boca Raton, FL.: CRC Press, Taylor and Francis (2015)
2. Chen, L.: Word Segmentation in Written Chinese (汉语书面语的分词问题——一个有关全民的信息化问题). Journal of Chinese Information Processing, No. 1, Vol. 10 (1996)
3. DeFrancis, J.: ABC Chinese-English Comprehensive Dictionary. Shanghai: Hanyu Dacidian Press (2003)

4. Hanban: New HSK Word List (新 汉 语 水 平 考 试 词 汇 表). Beijing: Hanban. http://www.chinesetest.cn/godownload.do (2012)

5. Internet World Stats: "INTERNET WORLD USERS BY LANGUAGE, http://www.internetworldstats.com/stats7.htm (2016)

6. Kleeman, J. and Yu, H.: The Oxford Chinese Dictionary. Oxford: Oxford University Press (2010)

7. Li, K. (李 克 强): Government Work Report 2016 (政 府 工 作 报 告 2016). http://www.gov.cn/guowuyuan/2016-03/05/content_5049372.htm (2016)

8. Linguistic Institute of the Chinese Academy of Social Sciences: Xinhua Dictionary (Xinhua Zidian, 新华字典), Beijing: The Commercial Press (2011)

9. Linguistic Institute of the Chinese Academy of Social Sciences: Contemporary Chinese Dictionary (Xiandai Hanyu Cidian, 现代汉语词典), Beijing: The Commercial Press (2012)

10. MDBG: CC-CEDICT Chinese to English Dictionary, URL: http://www.mdbg.net/chindict/chindict.php?page=cedict . (Downloaded on January 23, 2016)

11. National Language Commission of China (国家语委): The List of Frequently-used Words of Modern Chinese (现代汉语常用词表（草案）). Beijing: the Commercial Press (2008)

12. National Language Commission of China: Language Situation in China: 2014. (中国语言生活状况报告 (2014)). Beijing: Commercial Press (2014)

13. National Taiwan Normal University: New Word List of TOCFL (Test of Chinese as a Foreign Language). http://www.tw.org/tocfl/ (2013)

14. Yahoo: 教宗訪希臘難民營　帶走12敘利亞難民返梵蒂岡. https://hk.news.yahoo.com/教宗訪希臘難民營-帶走12敘利亞難民返梵蒂岡-033300795.html (2016).

15. Zhang, K., Sun, M.: Unified Framework of Performing Chinese Word Segmentation and Part-of-Speech Tagging. China Communications (中国通信), Vol.9(03), pp.1-9 (2012)

16. Zhang, X..: Written Chinese Word-Segmentation Revisited: Ten advantages of word-segmented writing. Journal of Chinese Information Processing (中文信息学报), No. 3, Vol. 12, pp. 57 - 63 (1998)

17. Zhang, X. and Li X.: Computer-Assisted Revision of the YES-CEDICT Chinese Dictionary (《一二三汉英大词典》的计算机辅助修订). To appear In Li, X. and Xu, J. eds. (李晓琪，徐娟主编), Digital Teaching of Chinese Language 2016 (数字化汉语教学 2016). Beijing: Tsinghua University Press (2016)

18. Zhang, X., Li, X. and Lun, C. (editors): The YES-CEDICT Chinese Dictionary. The Journal of Modernization of Chinese Language Education, Vol.4, No.1, June, 2015. (http://xuebao.eblcu.com/) (2015)