Lifetime Achievement Award

Translating Today into Tomorrow

Sheng Li
Harbin Institute of Technology

Good afternoon, ladies and gentlemen. I am standing here, grateful, excited, and proud. I see so many friends, my colleagues, students, and many more researchers in this room. I see that the work we started 50 years ago is now flourishing and is embedded in people's everyday lives. I see for the first time that the ACL conference is held here in Beijing, China. And I am deeply honored to be awarded the Lifetime Achievement Award of 2015.

I want to thank the ACL for giving me the Lifetime Achievement Award of 2015. It is the appreciation of not only my work, but also of the work that my fellow researchers, my colleagues, and my students have done through all these years. It is an honor for all of us. As a veteran of NLP research, I am fortunate to witness and be a part of its long yet inspiring journey in China. So today, to everyone here, my friends, colleagues, and students, either well-known scientists or young researchers: I'd like to share my experience and thoughts with you.

1. Early Machine Translation in China

The history of machine translation (MT) in China dates back to 1956. At that time the new country had immense construction projects to recover what had been ruined in the war. However, the government precisely recognized the significance of machine translation, and started to explore this area, as the fourth country following the United States, the United Kingdom, and the Soviet Union. In 1959, Russian–Chinese machine translation was demonstrated on a Type-104 general-purpose computer made in China. This first MT system had a dictionary of 2,030 entries, and 29 groups of rules for lexical analysis. Programmed by machine instructions, the system was able to translate nine different types of sentences. It used punched tape as the input, and the output was a special kind of code for Chinese characters, since there was no Chinese character output device at the time. As the pioneer in Chinese MT, the system touched the issues of word sense disambiguation, word reordering, and proposed the idea of predicate-focused sentence analysis and pivot language for multilingual translation. In the same year, machine translation research at the Harbin Institute of Technology (HIT) was started by Prof. Zhen Wang (and later Prof. Kaizhu Wang), focusing on the Russian–Chinese MT group. The pursuit for MT has never halted after these forerunners.

2. The CEMT Series

In 1960, I was admitted to HIT. Five years later, I graduated and became a faculty member in the computer department of HIT, which was probably the first computer discipline among Chinese universities. I started my research, however, not from machine translation but from information retrieval (IR). I was fully occupied by how to effectively store books and documents on computers, and then retrieve them quickly and accurately. The start of my research in MT was incidentally caused by IR problems.

At that time, Ming Zhou was my Ph.D. student. He is now the principal researcher of Natural Language Computing at Microsoft Research Asia (MSRA), and many of you may be acquainted with him. In 1985, at the beginning of his graduate study, he was aiming to address the topic of word extraction for Chinese documents to boost IR performance. For an exhaustive survey, Ming went to Beijing from Harbin alone, and buried himself at the National Library for over a month. He came back disappointed, finding that the related work was some language-dependent solutions for English. Actually, many research directions encountered this problem at that time. That's why Ming and I decided to develop an MT system through which we could first translate Chinese materials into English, so as to take advantage of the solutions proposed for English, and finally translate the results back into Chinese, if necessary.

In those years, the translation from Chinese to other foreign languages was less studied in China. Everything was hard in the beginning. We had to build everything from scratch, such as collecting and inputting each entry of the translation dictionary. Fortunately, we were not alone. I came to know many peer scholars, including Prof. Weitian Wu, Zhiwei Feng, Prof. Zhendong Dong, Prof. Shiwen Yu, and Prof. Changning Huang, as well as Dr. Zhaoxiong Chen. Although we didn't work together, we could always learn from each other and inspire each other in MT research.

After three years' effort, we accomplished a rule-based MT system named CEMT-I (Li et al. 1988). It ran on an IBM PC XT[1] and was capable of translating eight kinds of Chinese sentence patterns with fewer than one thousand rules. It had a dictionary of 30,000 Chinese-English entries. Simple or even crude as it now seems, it really encouraged every member of our team. After that, we developed CEMT-II (Zhou et al. 1990) and CEMT-III (Zhao, Li, and Zhang 1995) successively. The CEMT series seemed to have a special kind of magic. Almost all the students who participated in these projects devoted themselves to machine translation in their following careers, including Ming Zhou, Min Zhang, and Tiejun Zhao.

3. DEAR and BT863

Inspired by the success of the CEMT series, we also developed a computer-aided translation system called "DEAR." DEAR was put to market via a software chain store. Although it did not sell much, it was our first effort to commercialize the MT technology. I still remember how excited I was when I saw DEAR placed on the shelves for the first time. Today, it still reminds me that research work cannot just stay in the lab.

Also in the 1980s, China's NLP field was marked by a milestone event: the establishment of the Chinese Information Processing Society of China (CIPS). From then on, NLP researchers throughout the country have been connected and the academic exchange has been facilitated at the national scale. It was far beyond my imagination then that,

---

1 https://en.wikipedia.org/wiki/IBM_Personal_Computer_XT.

thirty years later, I would have the honor to be the president of this society, leading it to keep on contributing to the development of world-level NLP technology.

I usually regard the series of MT systems that we developed as a large family. In 1994, BT863 joined this family with some new features (Zhao, Li, and Wang 1995; Wang et al. 1997). First, BT863 was distinguished as a bi-direction translation between Chinese and English under a uniform architecture. Second, in addition to the rules, it was augmented with examples and templates learned from a corpus. Finally, this system is remembered for its top performance in the early national MT evaluation organized by the 863 High Tech Program of China.

4. Syntactic and Semantic Parsing

Time passed quickly. The rising of the Internet made communication more convenient, and our research was gradually connected with international peers. We concentrated on the mining and accumulation of bilingual and multilingual corpora. We explored how to integrate rule-based and example-based MT models under a unifying statistical framework. However, as more and more work was conducted, I found it more difficult to go deeper. I began to realize that translation problems cannot rely only on translation methods.

From word segmentation, morphology, word meaning, to named entity, syntax, and semantics, every step in this procedure affects the quality of translation. I remember an interesting story. One day, my student Wanxiang Che input his name into our machine translation system. The system literally translated his name into 'thousands of cars flying in the sky'. This was rated as the joke of the year in my lab, but the underlying problem is worth pondering.

Traditional Chinese medicine advocates the treatment of both symptoms and root causes. The same principle applies to MT research, in which models for word alignment, decoding, reordering, and so forth, can solve the surface problems of machine transla-tion, whereas understanding the word sense, sentence structure, and semantics is the solution to the fundamental problems. We therefore carried out research on syntactic analysis, including phrase-structure parsing and dependency parsing.

In those days, dependency parsing on Chinese was not widely studied. There was no well-accepted annotation or transformed standard. Therefore, we referred to a large number of linguistic studies, developed a Chinese syntactic dependency annotation standard, and annotated a 50,000-sentence Chinese syntactic dependency treebank on this basis. This is the largest Chinese dependency treebank available. Differently from those transformed from phrase-structure treebanks, our dependency structure uses native dependency structure, which can handle a large number of specific grammatical phenomena in dependency structures. This treebank has been released by the Linguistic Data Consortium (LDC) (Che, Li, and Liu 2012). We hope that more researchers can benefit from it.

Based on syntactic parsing, we hoped to further explore the semantic structure and the relationship of sentences. Therefore, we carried out research on semantic role labeling, and worked on the semantic role labeling methods based on the tree kernel, including the hybrid convolution tree kernel (Che et al. 2008) and the grammar-driven tree kernel (Zhang et al. 2007). In addition, we further broadened our mind and tried to analyze the semantics of Chinese directly. We proposed semantic dependency parsing tasks that directly establish semantic-level dependencies between content words, ignor-ing auxiliaries and prepositions. Meanwhile, we violated the tree structure constraints, allowing one word to depend on more than one parent node, so as to form semantic
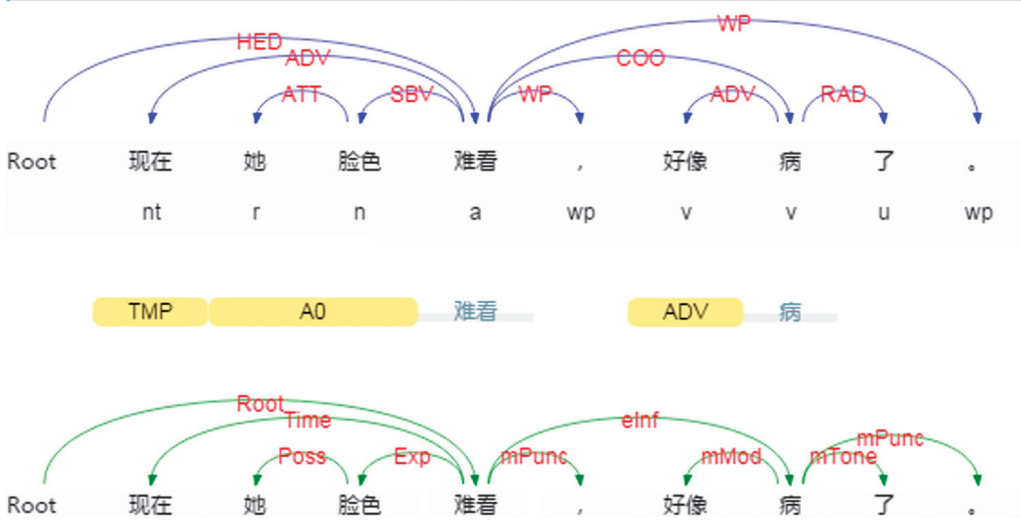
Figure 1
Example of syntactic dependency parsing, semantic role labeling, and semantic dependency parsing in LTP-Cloud.

dependency graph structures. At this point, the semantic dependency treebank that we have already labeled has reached more than 30,000 sentences. Much ongoing research is based on these data. Figure 1 shows an example of syntactic dependency parsing, semantic role labeling, and semantic dependency parsing for an input sentence "现在 / 她 / 脸色 / 难看 / ， / 好像 / 病了 / 。 [Now she looks terrible, seems to be sick]" .

5. LTP and LTP-Cloud

Every summer, HIT and MSRA would jointly organize a summer school for NLP research students. We invited domestic and foreign experts to give lectures to Chinese students engaged in this field. Because the summer school was free, students from all over the country came together every year, listening to lectures and conducting experiments. When I communicated with these students, I found that many of them came from labs that lacked fundamental NLP tools, such as word segmentors, part-of-speech taggers, and syntactic parsers. It would have been very difficult for them to implement their research ideas without these tools. I felt bad when I saw that. They are all students with dreams and innovative ideas. We must create a level playing field for all of them, I thought.

After coming back from the summer school, I met Ting Liu. He is a strong supporter of the idea of sharing. We decided to release an open-source NLP system: Language Technology Platform (LTP). This platform integrates several Chinese NLP basic technologies, including Chinese word segmentation, part-of-speech tagging, named entity recognition, dependency parsing, and semantic role labeling, which has made great contributions to the development of further applications.

In recent years, we realized that cloud computing and the mobile Internet have brought great opportunities and challenges to the NLP field. Therefore, we developed

LTP-cloud[2] in 2013, which provides accurate and fast NLP service via the Internet. Currently, the number of LTP-cloud registered users has exceeded 3,000, and most of them are NLP beginners. As I had wished, they no longer need to build an NLP basic processing system from scratch for their research. Every time I see the thank-you notes to the LTP and LTP-cloud in the acknowledgments of their papers, I am proud and grateful.

6. Machine Translation on the Internet

As more and more papers were published in top conferences and journals, our lab made a name in the academic world. Many people in the lab were satisfied, but I felt differently, since publishing papers should not be the major objective for research. New models and techniques should be applied to solve real-world problems and improve people's daily lives. Particularly since we have moved into the era of the Internet. Many new concepts and ideas have come into being, such as big data and cloud computing. In such a new era, machine translation research should no longer be restricted to the labs, running experiments on a small parallel corpus. Instead, it should embrace the Internet, and embrace big data. We paid great attention to the cooperation with IT and Internet companies. We established a joint lab with MSRA right after it was founded. After that, we have also established joint labs with other companies, like IBM, Baidu, and Tencent.

My student Haifeng Wang is the vice president of Baidu. He is in charge of NLP research and development, as well as Web search. We decided to collaborate in MT shortly after he joined Baidu, since Baidu can provide a huge platform for us to verify our ideas. Together with Tsinghua University, Zhejiang University, the Institute of Computing Technology, and the Institute of Automation of the Chinese Academy of Science, we successfully applied for an 863 project titled "Machine Translation on the Internet." All the members participating in this project have great passion for MT technologies and products.

Chinese people accept the principle that "取之于民，用之于民" [what is taken from the people should be used for the interests of the people]. Internet-based machine translation also follows this principle, which mines a large volume of translation data from the Internet, trains the translation model, and then provides high-quality services for Internet users. In our online translation system, taking Chinese–English translation, for example, there are hundreds of millions of parallel data for this language pair, which were filtered from billions of raw parallel data. We collected a large amount of data from hundreds of billions of Web pages. So I should say our MT service is actually built upon the whole Internet.

We have designed various mining models for these heterogeneous Internet data sources, including bilingual parallel pages, bilingual comparable pages, Web pages containing aligned sentence pairs, as well as plain texts containing entity and terminology translations. The mined translation data are filtered and refined. We set different updating frequencies for different Web sites, so as to guarantee that the latest data can be included. I often observe the mined translation data by myself, and I can find plenty of wonderful translations generated by ordinary Internet users. Their wisdom is perfectly integrated into the translation system. However, how to make use of such a big corpus?

---

2 http://www.ltp-cloud.com/demo/.

Figure 2
Examples of the Baidu online machine translation service.

This is a sweet annoyance. To handle big data, we have developed fast training and parallel decoding techniques in our project.

With such big data and frequent updates, even Internet buzzwords can be correctly translated. My students often post the so-called "magic translation" on microblogs. After the machine translation service came online, I began to realize that it would not only influence those Ph.D. students who are reading and writing research papers, or those businessmen who are studying materials from foreign countries. It also makes a huge difference to ordinary people's lives. Figure 2 shows some examples of Chinese–English machine translation from the Baidu online translation service,[3] which integrates the research work of the 863 project "Machine Translation on the Internet."

I once met a 50-year-old Chinese lady on a flight to Japan. She could not speak Japanese, but she had finally decided to marry her Japanese husband, with whom she had chatted online using machine translation. Another story comes from my neighbors, who are a couple my age. Their children have lived in Germany for years. The first
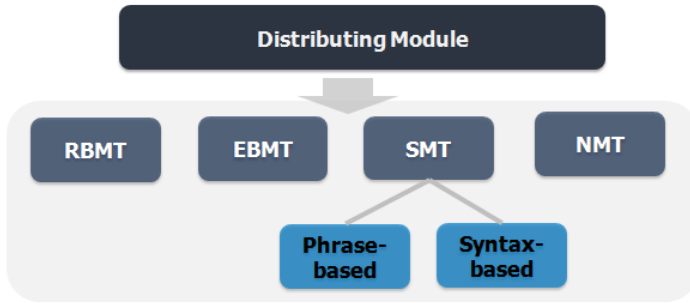
---

3 http://fanyi.baidu.com/.

Figure 3
Integration of MT models.

time the old couple met their grandson when their family came back to China, they were thrilled. However, on meeting their grandson, who can only speak German, they had no way to express their love, which made them sad. The grandma blamed herself and even wept when she was alone. With my recommendation, they started to use the online speech translation app in their smart phone. Now, they can finally talk to their grandson.

7. Integration of MT Models

I have been working in machine translation for several decades, going through almost all the streams of technologies, from rule-based MT (RBMT) models at the very beginning, example-based MT (EBMT) methods, to statistical MT (SMT) methods, as well as the research hotspot today—neural network machine translation (NMT). Actually, we tried the neural network–based models in NLP tasks, such as dialogue act analysis and word sense disambiguation, more than 15 years ago (Wang, Gao, and Li 1999a, 1999b). It is big data and computing power today that help neural network–based models significantly outperform traditional ones. I know that every method has its advantages and disadvantages. Although the new model and its methodology surpasses the old ones overall, it does not mean that the old methods are useless. There's an old saying in Chinese, "the silly bear keeps picking corn," which describes that when a bear is stealing corn from a peasant's field, it would always throw away the old one in its hand when it picked a new one; the silly bear would always end up with only one corn in his hand. I hoped that my team and I wouldn't become the "silly bears." Therefore, when we decided to develop an Internet MT system, we all agreed on the idea that we needed a hybrid approach, with which we could integrate all translation models and subsystems, on each of which we have all spent great effort. It is just like an orchestra, in which all instruments, such as piano, violin, cello, trumpet, and so on, are arranged perfectly together. Only in this way can the orchestra present a wonderful performance. As shown in Figure 3, in our MT system today, different models work together perfectly.

The rule-based method is used to translate expressions like date, time, and number. The example-based method is applied to translate buzzwords, especially the new emerging Internet expressions. On the other hand, those complicated long sentences are translated using the syntax-based statistical model, while those sentences that can be covered by a predefined vocabulary are translated with an NMT model. Finally, the

sentences left are all translated with a classical SMT model. The conductor of such an orchestra is a discriminative distributing module, which decides what subsystem an input sentence should be distributed to, based on a variety of statistical and linguistic features.

8. Translation for Resource-Poor Languages

Shortly after the release of Chinese–English and English–Chinese translation services, we also released translation services between Chinese and Japanese, Korean, and other daily-used foreign languages. However, with translation directions expanded, users' expectations for the translation between the resource-poor languages became higher and higher. Especially in recent years, China has been doing business more frequently with many countries, such as Thailand and Portugal, among others, and the destinations for Chinese tourists have become more diverse. One of my friends told me a story after he came back from a tour in Southeast Asia. He ordered three kinds of salads in a restaurant, since he did not understand or speak the local language. He could not communicate with the waiters or even read the menu. These incidents told us that solving translation problems for these resource-poor languages is urgent. Therefore, we have successively released translation services between Chinese and over 20 foreign languages. Now, we have covered languages in eight of the top ten destinations for Chinese tourists, and all the top ten foreign cities where Chinese tourists spend the most money.

On this basis, we took a further step. We built translation systems between any two languages using the pivot approach (Wang, Wu, and Liu 2006; Wu and Wang 2007). For resource-poor language pairs, we use English or Chinese as the pivot language. Translation models are trained for source-pivot and pivot-target, respectively, which are then combined to form the translation model from the source to the target language. Using this model, Baidu online translation services successfully realized pairwise translation between any two of 27 languages; in total, 702 translation directions.

9. MT Methodology for Other Areas

"他山之石，可以攻玉" [Stones from other hills may serve to polish the jade at hand]. This is a Chinese old saying from "《诗经》" [The Book of Songs], which was written 2,500 years ago. It suggests that one may benefit from other people's opinions and methods for their task. Machine translation technology is now a "stone from another hill," which has been used in many other areas. For instance, some researchers recast paraphrasing as a monolingual translation problem and use MT models to generate paraphrases of the input sentences (Zhao et al. 2009, 2010). There are also researchers who regard query reformulation as the translation from the original query to the rewritten one (Riezler and Liu 2010). However, what interests me the most is the encounter between translation technology and Chinese traditional culture. For example, MSRA uses the translation model to automatically generate couplets (Jiang and Zhou 2008), which are posted on the doors of every house during Chinese New Year. Baidu applies translation methods to compose poems. Given a picture and the first line of a poem, the system can generate another three lines of the poem that describe the content of the picture. In addition, I have heard recently that both Microsoft and Baidu have released their chatting robots, which are named Microsoft XiaoIce and Baidu Xiaodu, respectively. They both use translation techniques in the searching and generation of chatting responses. It is fair to say that machine translation has become more than a specific

method. Instead, it has evolved into a methodology and could make a contribution to other similar or related areas.

10. Conclusion

There is an ancient story in China called "愚公移山" [Yugong moves the mountain]. In the story, an old man called Yugong—meaning an unwise man—lived in a mountain area. He decided to build a road to the outside world by moving two huge mountains away. Other people all thought it was impossible and laughed at him. However, Yugong said to the people calmly: "Even if I die, I have children; and my children would have children in the future. As the mountain wouldn't grow, we would move the mountain away eventually." Today, when facing the ambitious goal of automatic high-quality machine translation, and even the whole NLP field, I cannot help thinking of Yugong's spirit. I have been, and I still am, trying to solve the questions and obstacles along the way. Even if one day I will no longer be able to keep exploring MT, I believe that the younger generations will keep on going until the dream of making a computer truly understand languages eventually comes true.

My friends, especially the young ones, to share what I have learned from my career, I'd like to say: Make yourself a good translation system: Input diligence today, and it will definitely translate into an amazing tomorrow!

References

Che, Wanxiang, Zhenghua Li, and Ting Liu. 2012. Chinese dependency treebank 1.0 LDC2012T05. *Web Download. Philadelphia: Linguistic Data Consortium, 2012.*

Che, Wanxiang, Min Zhang, AiTi Aw, ChewLim Tan, Ting Liu, and Sheng Li. 2008. Using a hybrid convolution tree kernel for semantic role labeling. *ACM Transactions on Asian Language Information Processing (TALIP)*, 7(4):13.

Jiang, Long and Ming Zhou. 2008. Generating Chinese couplets using a statistical MT approach. In *Proceedings of COLING*, pages 377–384, Manchester.

Li, Sheng, Ming Zhou, Miao Shi, and Weitian Wu. 1988. A Chinese–English machine translation system: CEMT-I. *Journal of China Society for Scientific and Technical Information*, 7(6):409–416.

Riezler, Stefan and Yi Liu. 2010. Query rewriting using monolingual statistical machine translation. *Computational Linguistics*, 36(3):569–582.

Wang, Haifeng, Wen Gao, and Sheng Li. 1999a. Dialog act analysis of spoken Chinese based on neural networks. *Chinese Journal of Computers*, 22(10):1014–1018.

Wang, Haifeng, Wen Gao, and Sheng Li. 1999b. Word sense disambiguation of spoken Chinese using neural network. *Journal of Software*, 10(12):1279–1283.

Wang, Haifeng, Sheng Li, Tiejun Zhao, Yan Yang, Endong Xun, and Min Zhang. 1997.

The research and implementation of a bilingual machine translation system (BT863) for Chinese and English. *Journal of the China Society for Scientific and Technical Information*, 16(5):360–369.

Wang, Haifeng, Hua Wu, and Zhanyi Liu. 2006. Word alignment for languages with scarce resources using bilingual corpora of other language pairs. In *Proceedings of COLING/ACL*, pages 874–881, Sydney.

Wu, Hua and Haifeng Wang. 2007. Pivot language approach for phrase-based statistical machine translation. In *Proceedings of the ACL*, pages 856–863, Prague.

Zhang, Min, Wanxiang Che, Aiti Aw, Chew Lim Tan, Guodong Zhou, Ting Liu, and Sheng Li. 2007. A grammar-driven convolution tree kernel for semantic role classification. In *Proceedings of the ACL*, pages 200–207, Prague.

Zhao, Shiqi, Xiang Lan, Ting Liu, and Sheng Li. 2009. Application-driven statistical paraphrase generation. In *Proceedings of the ACL/IJCNLP*, pages 834–842, Suntec.

Zhao, Shiqi, Haifeng Wang, Xiang Lan, and Ting Liu. 2010. Leveraging multiple MT engines for paraphrase generation. In *Proceedings of COLING*, pages 1326–1334, Beijing.

Zhao, Tiejun, Sheng Li, and Haifeng Wang. 1995. Pattern-based machine translation: Accomplishment of BT863 system. In

Proceedings of the NLPPRS, pages 371–376, Seoul.

Zhao, Tiejun, Sheng Li, and Min Zhang. 1995. CEMT-III: A fully automatic Chinese–English MT system. In *Proceedings of International Conference for Chinese Computing*, pages 8–14, Singapore.

Zhou, Ming, Sheng Li, Mingzeng Hu, and Shi Miao. 1990. An interactive Chinese–English machine translation system: CEMT-II. *Journal of the China Society for Scientific and Technical Information*, 9(2):151–154.