Chinese Sentence Compression: Corpus and Evaluation

Chunliang Zhang, Minghan Hu, Tong Xiao, Xue Jiang, Lixin Shi, Jingbo Zhu

Natural Language Lab, Northeastern University, Shenyang, China, 110819

{zhangcl, huminghan, xiaotong, zhujingbo}mail.neu.edu.cn;
{neujiangxue, shilixin}@163.com

Abstract. In this paper we present a first-ever manually-built Chinese sentence compression corpus. Based on this corpus, we develop a Chinese sentence compression system and study various measures for evaluation of Chinese sentence compression. We find that 1) using multi-references is very helpful for automatic evaluation in Chinese sentence compression; and 2) besides relational F1, some machine translation evaluation measures are correlated well with human judgments and thus are very promising for future use in this task.

Keywords. sentence compression, Chinese corpus, system evaluation

1 Introduction

Recent years have seen increasing interests in automatic sentence compression among the natural language processing researchers for a wide range of practical applications, such as text summarization, machine translation, and question answering. In general, the task of sentence compression can be described as creating a shorter form of a sentence while retaining the most important information and remaining grammaticality [8]. To date, many statistical models have been developed in sentence compression, showing continuous improvements on several tasks ([3-4], [7], [10-11], [18], [21-22]).

Naturally, as with any NLP technique, developing sentence compression systems relies on manually annotated corpora for training model parameters (in a supervised manner), system tuning, and evaluation of final results. However, the scarcity of such data restricts most work in English compression tasks (e.g., the Ziff-Davis corpus) and it is rare to see efforts in other languages.

In this paper we study the sentence compression problem for Chinese, one of the most popular languages other than English. We regard sentence compression as a task of identifying the selection word sequence of a sentence. In this way a compressed sentence is in principle a backbone of the original sentence and can be generated by removing all “unimportant” words. The contributions of this work are two-fold:

- We manually develop a Chinese sentence compression corpus consisting of 3,308 sentences from the Penn Chinese Treebank. For each sentence, there is at least one annotation. In addition, we provide three annotations for a sub-set of 563 sentences, which can be used as benchmark for evaluation of Chinese sentence compression systems. To the best of our knowledge, this is the first-ever manually annotated corpus for Chinese sentence compression.

- We study various evaluation metrics on the developed system for Chinese sentence compression. We find that 1) using multiple references is more helpful for automatic evaluation of the system performance than using the single-reference, as the strategy adoptedin previous studies; 2) four evaluation metrics, grammatical relations F1, mWER, mPER and GTM, have good co-relations with human judgments when used to measure the performance of the compression systems in terms of grammaticality and importance, and, therefore, are more desirable for automatic evaluation of Chinese sentence compression.

More importantly, our annotated corpus (as well as the annotation guideline, the automatic system output and human judgments) is accessible to public1 and can be used in further study and system development for this task. We believe that the developed corpus would motivate more studies on identifying the skeleton/main structure of Chinese sentences and would thus benefit many downstream NLP applications, such as machine translation and text summarization.

2 Related Work

Most previous work addresses the sentence compression task on English corpora. The most famous of these is the Ziff-Davis corpus [9], a collection of 1,067 sentences created automatically by matching sentences in a news article with sentences contained in its abstract. Yamangil and Nelken [20] collected a large-scale corpus of over 380,000 sentence pairs by mining the Wikipedia revision history of the articles and picking out those sentences with the record of word addition or deletion. But their work was based on an assumption that all the edits retain the core meaning of the sentence. There are two corpora manually created for English sentence compression [2], one is a 1,433-sentence dataset built from the British National Corpus and the American News Text Corpus, and the other is a 1,370-sentence dataset from the HUB-4 1996 English Broadcast News Corpus. However, to our knowledge, there is no such data in Chinese for sentence compression research.

3 Corpus Development

3.1 Data Selection

The original data in this work comes from the source-language side of the Penn Parallel Chinese-English Treebank (LDC2003E07). We choose this data set for annotation because all the sentences in the Penn Chinese Treebank (CTB) are of very good quality[19]. As these CTB sentences have been manually annotated with word segmentation, POS tags and syntactic structures, we believe they will be useful in studying the sentence compression problem on different conditions, e.g., comparison of the results obtained on gold-standard and automatic word segmentations/syntactic trees. Besides,

[1] http://202.118.18.77:8080/ChineseSentenceCompression/

our dataset in Chinese parallels with its English counterpart, and thus can be used in future studies of bilingual sentence compression or applying compression results in machine translation.

For convenience of annotation, we divide the selected dataset into 10 parts. Parts 1-8 consist of articles 001-270 and are with one annotation. Part 9 and Part 10 consist of articles 271-300 and articles 301-325 respectively, and both are with three annotations.

Table 1. The dataset used for annotation

Dataset	# Sentences	# Words	# Annotators
All (#1-3308)	3308	74312	1-3
Parts 1-8 (#1-2745)	2745	62868	1
Part 9 (#2746-3018)	273	5131	3
Part10 (#3019-3308)	290	6313	3

3.2 Annotation Guideline

This study focuses on Chinese sentence compression mainly for identifying the main structure of the Chinese sentence. We view sentence compression as a task of keeping the most important grammatically-motivated items of a sentence and removing all unimportant items. So in this work the result of sentence compression is in essential a grammatically-motivated skeleton of the input sentence.

In creating annotations for sentence compression, annotators are provided with the sentences only with word segmentation[2] and are required to compress the sentences by deleting the unimportant words while remaining sentence grammaticality.

Similar to the English counterpart, a Chinese sentence is composed of several constituents: the subject, the predicate, the object, the attributive, the adverbial, and the complement. The subject, the predicate, and the object are primary constituents, and the attributive, the adverbial, and the complement are secondary ones.

Original:
<晌午> 的 <太阳> || <火辣辣> 地 <烤> 着 <田野> 。
<noon>DE1<sun> || <fiery> DE2<scorch>ZHE <field>
attr. sub. || adv. pred. obj.
(The sun is scorching the field like fire at noon.)

Compressed:
<太阳> || <烤> 着 <田野> 。
<sun> || <scorch> ZHE <field>
sub. || pred. obj.
(The sun is scorching the field.)

Fig. 1. A demo of Chinese sentence compression with the sentence constituent analysis

[2] Word segmentation is a necessary step for most natural language processing tasks on Chinese for there is no delimiter between Chinese words. In this task, the information other than word segmentation is not available for annotators.

Basically, to achieve sentence compression, the first thing is to identify different constituents in a sentence, and then to remove the secondary constituents and retain the primary ones, because we believe the primary constituents constitute the structural backbone and carry the most valuable information in a sentence, and the secondary constituents just act as modifiers of one primary constituent and carry unimportant information. As shown in Fig.13, the annotators should first decompose the sentence into different constituents: the subject(sub.), the predicative(pred.), the object(obj.), the attributive(attr.), and the adverbial(adv.). Note that this Chinese sentence includes three auxiliary words: two structure-auxiliary words4 ‘的(DE1)’ and ‘地(DE2)’, which mark the preceding constituents as the attributive and the adverbial respectively, and one aspect-auxiliary word ‘着(ZHE)’5, which is attached to the preceding verb ‘烤(scorch)’ and acts as the indicator of the durative aspect for the verb. After the sentence constituents are identified, the sentence compression is done by deleting the attributive ‘晌午(noon)’, the adverbial ‘火辣辣(fiery)’ and their attached structural auxiliary words, ‘的(DE1)’ and ‘地(DE2)’ .

In practice, the word deletion is done at two levels: the word level and the phrase level. To save space, we list only a few critical annotation rules here[6].

At the word level, all the adjectives will be deleted if they modify a noun/noun phrase, as the phrase[7] ‘<高昂>的 <成本>(<high>DE1<cost>, high cost)’ is compressed as a noun ‘<成本>(cost)’ by deleting the adjective ‘<高昂>(high)’ and the structural auxiliary ‘的(DE1)’. Besides, the degree adverbs, such as ‘很(very)’ and ‘有点儿(a little)’, will be deleted if they modify an adjective, as the phrase ‘很 美 (very beautiful)’ is compressed as an adjective ‘美(beautiful)’.

Original:
<据 报道>， 朝鲜 代表团 ‖ 已经 抵达 北京 。
According to report, DPRK delegation ‖ already arrive Beijing.
(It is reported that the DPRK delegation has arrived at Beijing.)

Compressed:
代表团 ‖ 已经 抵达 北京 。
delegation ‖ already arrive Beijing.
(The delegation has arrived at Beijing.)

Fig. 2. A demo of the parenthesis deletion for Chinese sentence compression

[3] The double vertical lines‘‖’ in Fig. 1 shows the boundary between the subject and the predicate, the two primary constituents in a sentence.

[4] The Chinese auxiliary words, ‘的’, ‘地’, and ‘得’, are usually denoted as ‘DE1’, ‘DE2’ and ‘DE3’ respectively in analysis of the syntactic structure.

[5] The Chinese auxiliary words, ‘着(ZHE)’, ‘了(LE)’, ‘过(GUO)’, are attached to a verb to mark its aspect and tense.

[6] For detailed description of the annotation guideline, please refer to http://202.118.18.77:8080/ChineseSentenceCompression/.

[7] The ‘phrase’ used here, instead of the ‘sentence’ is for space-saving. It is by no means to compress a phrase in this work.

Table 2. A demo of some fundamental annotation rules for Chinese sentence compression

Comp. target	Example
adjectives	original: <*美丽的*> <蝴蝶>||<飞走> <了>。 (The *beautiful* butterfly || flew away.)
	compressed: <蝴蝶> || <飞走> <了>。 (The butterfly || flew away.)
degree adverbs	original: <这里的><景色> || <*真*美>。 (The scenery here is *really* beautiful.)
	compressed: <景色> || <美>。 (The scenery || is beautiful.)
noun phrases	original: <*中国 国家 主席 习近平*> || <将><于近日> <出访><俄罗斯>。 (*The Chinese president Xi Jinping* || will visit Russia in a few days.)
	compressed: <习近平>||<将><出访><俄罗斯>。 (Xi Jinping will visit Russia.)
prep. phrases	original: <老师> || <希望> <我们> <为了美好 的 明天> 而 <学习> 。 (The teacher || hopes we will study for a beautiful tomorrow.)
	compressed: <老师 <希望> <我们> <学习> 。 (The teacher || hopes we will study.)
parentheses	original: <在 中国 的 大 城市>，<尤其是 北京 和 上海>，<交通 堵塞> || <非常 严重>。 (In large cities in china, especially Beijing and Shanghai, traffic jam || is very serious.)
	compressed: <交通堵塞> <严重>。 (The traffic jam || is serious.)

At the phrase level, the compression rules are mainly concerned about the noun phrase and the prepositional phrase. For the noun phrase comprised of a succession of nouns, some nouns will be deleted if they modify the other nouns, as the noun phrase '美国 有线 新闻 网 记者(CNN correspondent)' is compressed as the noun '记者(correspondent)', for human annotators can easily distinguish the two units of the phrase, the unit of a proper noun, '美国 有线 新闻 网(CNN)', modifying the unit of a noun '记者(correspondent)'. Another kind of noun phrase is that it contains two coreferents, like the phrase '中国 国家 主席 习近平(Chinese president Xi Jinping)', where ''中国 国家 主席(Chinese president)', the job title, and '习近平(Xi Jinping)' , the person's name, corefer to each other. In such a case, the compression is done by deleting one of them (usually retaining the proper noun). For the prepositional phrase, it will be deleted when it functions as the adverbial in the sentence, as the phrase '为了美好 的 明天 而 学习(study for a beautiful tomorrow)' is compressed as '学习(study)' by deleting the prepositional phrase '为了美好 的 明天(for a beautiful tomorrow)'.

Besides the above rules, the parenthesis[8] in a sentence will be deleted during the compression, as in Fig. 2, the parenthesis '据 报道(it is reported)', which shows the source of information for the following statement, is deleted for sentence compres-

[8] The parenthesis refers to the elements in a sentence which functions as the explanatory or qualifying remarks and has no clear dependent relations with the other constituents of a sentence. The parenthesis is usually delimited with a comma if it locates at the beginning or the end of a sentence, or two commas if it locates in the middle of a sentence.

sion because it seems to be an independent element from the other parts of the sentence.

For better understanding of the fundamental rules discussed above for annotation of Chinese sentence compression, we list them in table 2 with examples.

3.3 Quality Control

Three annotators[9] participate in this task. To guarantee high annotation quality, we implement a two-phase process: phase 1 is a multi-round pilot annotation on small-size datasets for training and phase 2 is a formal annotation on the full size dataset.

Table 3. Statistics of three-round pilot annotation

Round	# Sentences	Compression Rate		Kappa
1	30	Human1	0.717	0.652
		Human2	0.685	
		Human3	0.657	
2	50	Human1	0.559	0.841
		Human2	0.581	
		Human3	0.566	
3	50	Human1	0.551	0.886
		Human2	0.557	
		Human3	0.535	

Table 4. Statistics of formal annotation on Parts 9-10

Dataset	# Sentences	Compression Rate		Kappa
Part9	273	Human1	0.543	0.885
		Human2	0.576	
		Human3	0.571	
Part10	290	Human1	0.512	0.860
		Human2	0.551	
		Human3	0.536	

For each round of annotation, we calculate the compression rate with each annotator's work and Fleiss' Kappa [6] for the inter-annotator agreement. After each round of annotation, the manual is revised based on our review of the inter-annotation inconsistencies and discussion about the ambiguous cases. Only after Fleiss' Kappa indicates the inter-annotator agreement is satisfactory and remains stable will the pilot annotation stop and the formal-run annotation start.

During the phase of formal annotation, two annotators work on Parts 1-8 (2,745 sentences), providing one reference result for each source sentence. To further check the inter-annotator consistency, the three annotators work on Part 9 and Part 10 respectively, and thus each sentence in these two datasets has three reference compression results.

---

[9] All three of the annotators are Chinese natives and have received considerable training in linguistics, particularly in syntax.

4 Evaluation and Analysis

Using the corpus presented above, we develop a Chinese sentence compression system and study various evaluation methods for this task.

4.1 Automatic Sentence Compression

We use Tree Transducer Toolkit (T3)[10] to build a Chinese sentence compression system. T3 is a tree-to-tree transduction model based on synchronous tree-substitution grammars (STSGs), which achieves state-of-the-art performance in the English sentence compression tasks [4]. To enable T3 to perform on the Chinese data, we modify the head-finding rules according to the Chinese head rules described in [1]. We use the data of Parts 1-8 as the training set, Part 9 as tuning set and Part 10 as the test set. To obtain the *n*-gram feature, we train a tri-gram language model on the Xinhua and AFP Portions of the GIGAWORD Chinese corpus. Since T3 requires CTB-style trees, we use the Berkeley parser[11] to parse all the sentences[12]. By default we choose the asymmetric hamming distance loss function for the large margin training of the system.

4.2 Evaluation Metrics

Like in English sentence compression tasks, we choose grammatical relation F1 as one of the evaluation metrics, which allows us to measure the semantic aspects of summarization quality in terms of grammatical-functional information [14]. We use the ZPar dependency parser[13][23] to extract Chinese grammatical relations for all the sentences in the test set and gold references.

In principle, the sentence compression evaluation is to compute the errors of a sentence against its gold reference(s), which is similar to the evaluation of MT systems, especially when no paraphrasing is performed in compression as we do in this work. Therefore, we adopt additional MT evaluation metrics in our experiment.Specifically we choose three *n*-gram and similarity-based metrics, BLEU[13], NIST[5] and GTM[17], which are very popular in automatic evaluation methods in MT. Besides, we use three Levenshtein distance based metrics, mWER[12], mPER[16], and TER[15], which regard the evaluation problem as pairwise string alignment between the output string and the gold reference[14].

To study the correlation between the automatic evaluation measures and human judgments, we also conduct human evaluation on the same data. Judges are required to separately rate along a 5-point scale how much information the compressed sen-

---

[10] http://staffwww.dcs.shef.ac.uk/people/T.Cohn/t3/

[11] http://code.google.com/p/berkeleyparser/downloads/list

[12] It would be interesting to compare the results of using automatic parsers and gold parse trees. We leave it for our future work.

[13] http://www.sutd.edu.sg/cmsresource/faculty/yuezhang/zpar.html

[14] Not that, for grammatical relation F1, BLEU, NIST and GTM, larger values reflect better translation quality. For mWER, mPER and TER, smaller means better.

tence retains against the source sentence (i.e., *importance*) and how grammatical the compression is without the presence of the source sentence (i.e., *grammaticality*).

Table 5. Evaluation results with different metrics and different number of references

Metrics	Ref1	Ref2	Ref3	Ref1-3
NIST	7.423	6.889	7.017	7.938
BLEU	0.521	0.508	0.498	0.641
GTM	0.694	0.700	0.685	0.739
mWER	0.516	0.497	0.520	0.445
mPER	0.471	0.457	0.459	0.407
TER	0.495	0.484	0.506	0.418
Relational F1	0.574	0.587	0.547	0.652

Table 6. Human evaluation results of different system outputs

Entry	Com.R	Importance	Grammaticality
Output 1	0.400	3.803	4.228
Output 2	0.341	3.438	4.017
Output 3	0.631	4.252	4.566
Output 4	0.424	3.700	4.183
Output 5	0.427	3.679	4.152

4.3 Results and Analysis

Table 6 shows the automatic evaluation results on the test set with single and multiple references. We see, first of all, that the results of each reference are not very stable and show irregular variance by different measures. We attribute this to the ambiguity of sentence compression tasks, that is, even though annotators can get agreement in most cases, there exists some cases with more than one correct answer. This explanation is further confirmed when we switch to the multi-reference evaluation. By the 3-reference result, all measures show significant different scores (or better performance) with the single-reference counterparts, indicating that the sentence compression task has some natural ambiguities which cannot be eliminated, even for well-trained native language annotators. Therefore, for reliable estimation of the compression system performance, it is necessary to conduct evaluation with more than one reference. This finding is actually somewhat similar to that in machine translation where correct translations are plenty and the evaluation against a single reference is very unreliable [13].

We then examine how the automatic measures correlate with human judgments. To conduct evaluation on diverse compression results, we generate five outputs with different loss functions used in the T3 toolkit[15]. 7 Chinese native-language judges participate in the evaluation and score each system output by the rating schema presented in Section 4.2. Table 6 shows that compression rate is an important factor for a

[15] The 5 outputs correspond with loss function 2, 8, 10, 16 and 17 respectively.

successful Chinese sentence compression system. For example, the best result (output 3) is achieved when the compression rate is closest to those of the references, while the worst result (output 2) corresponds with a compression rate that is the farthest from the reference rates as shown in Table 4.

Table 7. Correlation coefficients between automatic measures and human judgments

Entry	NIST	BLEU	GTM	mWER	mPER	TER	Relational F1
Importance	0.843	0.807	0.870	-0.887	-0.884	-0.718	0.896
Grammaticality	0.827	0.798	0.878	-0.904	-0.900	-0.732	0.905

Finally we plot the automatic measures as functions of the human evaluation scores[16]. As shown in Fig. 3, most of these measures correlate well with the human judgments on outputs 1, 2, 3 and 5. However, they cannot distinguish well between output 1 and output 4 which are quite close in human evaluation. This phenomenon reflects a limited ability of current automatic metrics in prediction on similar compression results. Furthermore, we use the Pearson Correlation Coefficients to estimate the correlation degree. Table 7 shows that relational F1 correlates best with judges (around 0.90), which agrees with the observation seen in the English tasks[3]. More interestingly, it is observed that GTM, mWER and mPER obtain very good correlation scores (absolute value > 0.87), followed by BLEU and NIST (absolute value > 0.79). These results indicate a very promising application of MT evaluation methods in Chinese sentence compression tasks.

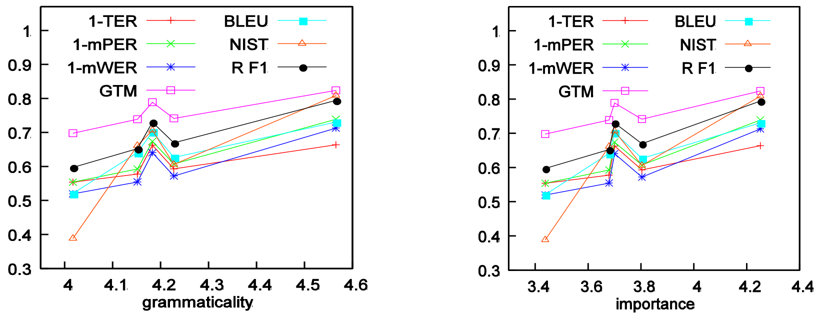

Fig. 3. Automatic measures of the sentence compression results against human judgments

5 Conclusion and Future Work

We have presented a first-ever manually-built Chinese sentence compression corpus. By using this corpus, we develop an automatic sentence compression system and study various evaluation methods on this task. We find that 1) using multiple references is necessary for automatic evaluation; and 2) besides relational F1, some MT

16 As mWER, mPER and TER have negative correlations with human evaluation scores, we use 1-mWER, 1-mPER and 1-TER as functions for a clear presentation. Also, the NIST score is normalized with a factor of 12 to fit it into the range of [0, 1].

evaluation measures are also well correlates with human judgments, and are very promising for the evaluation of sentence compression systems.

In the future we would like to enlarge this Chinese sentence compression corpus by annotating the other parts of the CTB data and apply the corpus to some NLP tasks like machine translation.

Acknowledgements

This work was supported in part by the National Science Foundation of China (61073140; 61272376), Specialized Research Fund for the Doctoral Program of Higher Education (20100042110031) and the Fundamental Research Funds for the Central Universities (N100204002).

We would like to thank Dirk Hovy for providing his kappa calculator. We also thank Yue Zhang for his assistance with the ZPar statistical parser. Finally we thank Matt Snover for assisting us to use the TER metric in our Chinese sentence compression evaluation work.

References

1. Bikel, D.M.: Intricacies of Collins' Parsing Model. In: Computational Linguistics, 30(4): 479-511(2004).
2. Clarke, J. and Lapata, M.: Global inference for sentence compression: An integer linear programming approach. In: Journal of Artificial Intelligence Research, 1(31):399–429(2008)
3. Clarke, J. and Lapata, M: Models for sentence compression: A comparison across domains, training requirements and evaluation measures. In: Proceedings of ACL-COLING, pp. 377-384(2006b)
4. Cohn, T. and Lapata, M.: Sentence compression beyond word deletion. In: Proceedings of the 22nd COLING, pp. 137-144(2009).
5. Doddington, G.: Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. In: Proceeding of HLT '02 Proceedings of the second international conference on Human Language Technology Research, pp. 138-145(2002)
6. Fleiss, J. L.: Measuring nominal scale agreement among many raters. In: Psychological Bulletin, 76(5): 378–382 (1971)
7. Galley, M. and McKeown, K.R.: Lexicalized Markov Grammars for Sentence Compression. In: Proceedings of HLT-NAACL, pp. 180–187 (2007)
8. Jing, H.: Sentence Reduction for automatic summarization. In: Proceedings of ANLP, pp. 310-315(2000)
9. Knight, K. and Marcu, D.: Statistical-based summarization-step one: sentence compression. In: Proceedings of AAAI-2000, pp. 703–710(2000)
10. Knight, K. and Marcu, D.: Summarization beyond sentence extraction: a probabilistic approach to sentence compression. In: Artificial Intelligence. 139(1):91-107(2002)
11. McDonald, R.: Discriminative sentence compression with soft syntactic constraints. In: proceedings of EACL, pp. 297-304(2006)

12. Nie ßen, S., Och, F. J., Leusch, G., and Ney, H.: An evaluation tool for machine translation: Fast evaluation for machine translation research. In: proceedings of the Second International conference on Language Resources and Evaluation (LREC), pp. 39–45 (2000)
13. Papineni, K., Roukos, S., Ward, T. and Zhu, W.J.: BLEU: a Method for Automatic Evaluation of Machine Translation. In: Proceedings of 40th Annual Meeting of ACL, pp. 311-318(2002)
14. Riezler, S., King, T.H., Crouch, R. and Zaenen, A.: Statistical sentence condensation using ambiguity packing and stochastic disambiguation methods for lexical-functional grammar. In *Proceedings of HLT-NAACL*, pp. 118-125(2003)
15. Snover, M., Dorr, B., Schwartz, R., Micciulla, L., and Makhoul, J.: A Study of Translation Edit Rate with Targeted Human Annotation. In: Proceedings of Association for Machine Translation in the Americas, pp. 223-231(2006)
16. Tillmann, C., Vogel, S., Ney, H., Zubiaga, A. and Sawaf, H.: Accelerated DP based search for statistical translation. In: Proceedings of European Conference on Speech Communication and Technology, pp. 2667–2670 (1997)
17. Turian, J. P., Shen, L., and Melamed, I. D.: Evaluation of Machine Translation and its Evaluation. In: Proceedings of MT Summit IX, pp. 386-393 (2003)
18. Turner, J. and Charniak, E.: Supervised and unsupervised learning for sentence compression. In: proceedings of 43th Annual Meeting of ACL, pp. 290-297(2005)
19. Xue, N., Xia, F., Chiou, F., Palmer, M.: The Penn Chinese TreeBank: Phrase structure annotation of a large corpus. In: Natural Language Engineering, 11(2): 207-238(2005)
20. Yamangil, E. and Nelken, R.: Mining Wikipedia revision histories for improving sentence compression. In: Proceedings of 46th Annual Meeting of ACL, pp. 137-140(2008)
21. Yamangil, E. and Shieber, S.M.: Bayesian synchronous tree-substitution grammar induction and its application to sentence compression. In: Proceedings of 48th Annual Meeting of ACL, pp. 934-947(2010)
22. Yoshikawa, K., Iida, R., Hirao, T. and Okumura, M.: Sentence Compression with Semantic Role Constraints. In: Proceedings of 50th Annual Meeting of ACL, pp. 349-353(2012)
23. Zhang, Y. and Clark, S.: Syntactic Processing Using the Generalized Perceptron and Beam Search. In: Computational Linguistics, 37(1), pp. 105-151(2011)