A Parse-and-Trim Approach with Information Significance
for Chinese Sentence Compression

Wei Xu   Ralph Grishman
Computer Science Department
New York University
New York, NY, 10003, USA
`{xuwei,grishman}@cs.nyu.edu`

Abstract

In this paper, we propose an event-based approach for Chinese sentence compression without using any training corpus. We enhance the linguistically-motivated heuristics by exploiting event word significance and event information density. This is shown to improve the preservation of important information and the tolerance of POS and parsing errors, which are more common in Chinese than English. The heuristics are only required to determine possibly removable constituents instead of selecting specific constituents for removal, and thus are easier to develop and port to other languages and domains. The experimental results show that around 72% of our automatic compressions are grammatically and semantically correct, preserving around 69% of the most important information on average.

1   Introduction

The goal of sentence compression is to shorten sentences while preserving their grammaticality and important information. It has recently attracted much attention because of its wide range of applications, especially in summarization (Jing, 2000) and headline generation (which can be viewed as summarization with very short length requirement). Sentence compression can improve extractive summarization in coherence and amount of information expressed within a fixed length.

An ideal sentence compression will include complex paraphrasing operations, such as word deletion, substitution, insertion, and reordering. In this paper, we focus on the simpler instantiation of sentence simplification, namely word deletion, which has been proved a success in the literature (Knight and Marcu, 2002; Dorr et al, 2003; Clarke and Lapata, 2006).

In this paper, we present our technique for Chinese sentence compression without the need for a sentence/compression parallel corpus. We combine linguistically-motivated heuristics and word significance scoring together to trim the parse tree, and rank candidate compressions according to event information density. In contrast to probabilistic methods, the heuristics are more likely to produce grammatical and fluent compressed sentences. We reduce the difficulty and linguistic skills required for composing heuristics by only requiring these heuristics to identify possibly removable constituents instead of selecting specific constituents for removal. The word significance helps to preserve informative constituents and overcome some POS and parsing errors. In particular, we seek to assess the event information during the compression process, according to the previous successes in event-based summarization (Li et al, 2006) and a new event-oriented 5W summarization task (Parton et al, 2009).

The next section presents previous approaches to sentence compression. In section 3, we describe our system with three modules, viz. linguistically-motivated heuristics, word significance scoring and candidate compression selection. We also develop a heuristics-only approach for comparison. In section 4, we evaluate the compressions in terms of grammaticality, infor-

mativeness and compression rate. Finally, Section 5 concludes this paper and discusses directions of future work.

2 Previous Work

Most previous studies relied on a parallel corpus to learn the correspondences between original and compressed sentences. Typically sentences are represented by features derived from parsing results, and used to learn the transformation rules or estimate the parameters in the score function of a possible compression. A variety of models have been developed, including but not limited to the noisy-channel model (Knight and Marcu, 2002; Galley and McKeown, 2007), the decision-tree model (Knight and Marcu, 2002), support vector machines (Nguyen et al, 2004) and large-margin learning (McDonald, 2006; Cohn and Lapata 2007).

Approaches which do not employ parallel corpora are less popular, even though the parallel sentence/compression corpora are not as easy to obtain as multilingual corpora for machine translation. Only a few studies have been done requiring no or minimal training corpora (Dorr et al, 2003; Hori and Furui, 2004; Turner and Charniak, 2005). The scarcity of parallel corpora also constrains the development in languages other than English. To the best of our knowledge, no study has been done on Chinese sentence compression.

An algorithm making limited use of training corpora was proposed originally by Hori and Furui (2004) for spoken text in Japanese, and later modified by Clarke and Lapata (2006) for English text. Their model searches for the compression with highest score according to the significance of each word, the existence of Subject-Verb-Object structures and the language model probability of the resulting word combination. The weight factors to balance the three measurements are experimentally optimized by a parallel corpus or estimated by experience.

Turner and Charniak (2005) present semi-supervised and unsupervised variants of the noisy channel model. They approximate the rules of compression from a non-parallel corpus (e.g. the Penn Treebank) based on probabilistic context free grammar derivation.

Our approach is most similar to the Hedge Trimmer for English headline generation (Dorr et al, 2003), in which linguistically-motivated heuristics are used to trim the parse tree. This method removes low content components in a preset order until the desired length requirement is reached. It reduces the risk of deleting subordinate clauses and prepositional phrases by delaying these operations until no other rules can be applied. This fixed order of applying rules limits the flexibility and capability for preserving informative constituents during deletions. It is likely to fail by producing a grammatical but semantically useless compressed sentence. Another major drawback is that it requires considerable linguistic skill to produce proper rules in a proper order.

3 Algorithms for Sentence Compression

Our system takes the output of a Chinese Treebank-style syntactic parser (Huang and Harper, 2009) as input and performs tree trimming operations to obtain compression. We propose and compare two approaches. One uses only linguistically-motivated heuristics to delete words and gets the compression result directly. The other one uses heuristics to determine which nodes in the parse tree are potentially removable. Then all removable nodes are deleted one by one according to their significance weights to generate a series of candidate compressions. Finally, the best compression is selected based on sentence length and informativeness criteria.

3.1 Linguistically-motivated Heuristics

This module aims to identify the nodes in the parse tree which may be removed without severe loss in grammaticality and information. Based on an analysis of the Penn Treebank corpus and human-produced compression, we decided that the following parse constituents are potential low content units.

Set 0 – basic:
- Parenthetical elements
- Adverbs except negative, some temporal and degree adverbs
- Adjectives except when the modified noun consists of only one character
- DNPs (which are formed by various phrasal categories plus "的" and appear as modifiers of NP in Chinese)
- DVPs (which are formed by various phrasal categories plus "地" in Chinese, and appear as modifiers of VP in Chinese)
- All nodes in noun coordination phrases except the first noun

Set 1 – fixed:
- All children of NP nodes except temporal nouns and proper nouns and the last noun word
- All simple clauses (IP) except the first one, if the sentence consists of more than one IP
- Prepositional phrases except those that may contain location or date information, according to a hand-made list of prepositions

Set 2 – flexible:
- All nodes in verb coordination phrases except the first one.
- Relative clauses
- Appositive clauses
- All prepositional phrases
- All children of NP nodes except the last noun word
- All simple clauses, if the sentence consists of more than one IP (at least one clause is required to be preserved in later trimming)

Set 0 lists all the fundamental constituents that may be removed and is used in both approaches. Set 1 and Set 2 are designed to handle more complex constituents for the two approaches respectively.

The heuristics-only approach exploits Set 0 and Set 1. It can be viewed as the Chinese version of Hedge Trimmer (Dorr et al, 2003), but differs in the following ways:
1) Chinese has different language constructions and grammar from English.
2) We eliminate the strict compression length constraint in order to yield more natural compressions with varying length.
3) We do not remove time expressions on purpose to benefit further applications, such as event extraction.

The heuristics-only approach deletes low content units mechanically while preserving syntactic correctness, as long as parsing is accurate. Our preliminary experiments showed that the heuristics in Set 0 and Set 1 can generate a comparatively satisfying compression, but is sensitive to part-of-speech and parsing errors, e.g. the proper noun "现代 (Hyundai)" as motor company is tagged as an adjective (shown in Figure 1) and thus removed since its literal meaning is "现代(modern)". Moreover, the rules in Set 1 reduce the sentence length in a gross manner, risking

serious information or grammaticality loss. For example, the first clause may not be a complete grammatical sentence, and is not always the most important clause in the sentence though that is usually the case. We also want to point out that the heuristics tend to reduce the sentence length and preserve the grammar by removing most of the modifiers, even though modifiers may contain a lot of important information.

To address the above problems of heuristics, we exploit word significance to measure the importance of each constituent. Set 2 was created to work with Set 0 to identify removable low content units. The heuristics in this approach are used only to detect all possible candidates for deletion and thus are more general and easier to create than Set 1. For instance, we do not need to carefully determine which kinds of prepositional phrases are safe or dangerous to delete but instead mark all of them as potentially removable.

The actual word deletion is performed later by a compression generation and selection module, taking word significance and compression rate into consideration. The heuristics in Set 2 are able to cover more risky constituents than Set 1, e.g. clauses and parallel structures, since the risk will be controlled by the later processes.

```
( (IP
    (NP
        (*NP (NR 韩国))              South Korean
        (#*ADJP (JJ 现代))           Hyundai
        (NP
            (#*NN 汽车)              motor
            (NN 公司)))              company
    (VP (VC 是)                      is
        (NP (#*DNP (NP (NR 沃尔沃))   Volvo
                (DEG 的))            's
            (#*ADJP (JJ 潜在))        potential
            (NP (NN 买家))))          buyer
    (PU .)))
```
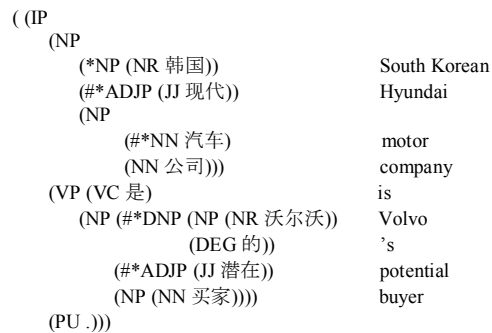
Figure 1. Parse tree trimming by heuristics (#: nodes trimmed out by Set0 & Set1; *: nodes labeled as removable by Set0 & Set2.)

Figure 1 shows an example of applying heuristics to the parse tree of the sentence "韩国现代汽车公司是沃尔沃的潜在买家" (The South Korean Hyundai Motor Company is a potential buyer of Volvo.). The heuristics-only approach produces "韩国公司是买家" (The South Korean company is a buyer.), which is grammatical but semantically meaningless. We will see how word significance and information density scoring produce a better compression in section 3.3.

3.2 Event-based Word Significance

Based on our observations, a human-compressed sentence primarily describes an event or a set of relevant events and contains a large proportion of named entities, especially in the news article domain. Similar to event-based summarization (Li et al, 2006), we consider only the event terms, namely verbs and nouns, with a preference for proper nouns.

The word significance score $I_j(w_i)$ indicates how important a word w_i is to a document j. It is a tf-idf weighting scheme with additional weight for proper nouns:

$$I_j(w_i) = \begin{cases} tf_{ij} \times idf_i, & \text{if } w_i \text{ is verb or common noun} \\ tf_{ij} \times idf_i + \omega, & \text{if } w_i \text{ is proper noun} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where
w_i : a word in the sentence of document j
tf_{ij} : term frequency of w_i in document j
idf_i : inverse document frequency of w_i
ω : additional weight for proper noun.

The nodes in the parse tree are then weighted by the word significance for leaves or the sum of the children's weights for internal nodes. The weighting depends on the word itself regardless of its part-of-speech tags in order to overcome some part-of-speech errors.

3.3 Compression Generation and Selection

In this module, we first apply a greedy algorithm to trim the weighted parse tree to obtain a series of candidate compressions. Recall that the heuristics Set 0 and 2 have provided the removability judgment for each node in the tree. The parse tree trimming algorithm is as follows:

1) remove one node with the lowest weight and get a candidate compressed sentence
2) update the weights of all ancestors of the removed node
3) repeat until no node is removable

The selection among candidate compressions is a tradeoff between sentence length and amount of information. Inspired by headlines in news articles, most of which contain a large proportion of named entities, we create an information density measurement $D(s_k)$ for sentence s_k to select the best compression:

$$D(s_k) = \frac{\sum\limits_{w_i \in P} I(w_i)}{L(s_k)} \quad (2)$$

where
P : the set of words whose significance scores are larger than ω in (1)
$I(w_i)$: the significance score of word w_i
$L(s_k)$: the length of sentence in characters

Table 1 shows the effectiveness of information density to select a proper compression with a balance between length and meaningfulness. Table 1 lists all candidate compressions in sequence generated from the parse tree in Figure 1. The words in bold are considered in information density. The underlined compression is picked as final output as "韩国现代公司是沃尔沃的买家" (The South Korean Hyundai company is a buyer of Volvo.), which makes more sense than the one produced by heuristics-only approach as "韩国公司是买家" (The South Korean company is a buyer.). In our approach, "现代(Hyundai)" tagged as adjective and "沃尔沃的(Volvo's)" as a modifier to buyer are preserved successfully.

D(s)	Sentence
0.254	**韩国现代**汽车公司是**沃尔沃**的潜在买家. The South **Korean Hyundai** Motor Company is a potential buyer of **Volvo**.
0.288	韩国现代汽车公司是沃尔沃的买家. The South Korean Hyundai Motor Company is a buyer of Volvo.
0.332	韩国现代公司是沃尔沃的买家. The South Korean Hyundai Company is a buyer of Volvo.
0.282	韩国公司是沃尔沃的买家. The South Korean company is a buyer of Volvo.
0.209	公司是沃尔沃的买家. The company is a buyer of Volvo.
0.0	公司是买家. The company is a buyer.

Table 1. Compression generation and selection for the sentence in Figure 1

The compression with highest information density is chosen as system output. To achieve a better compression rate and avoids overly condensed sentences (i.e. very short sentences with only a proper noun), we further constrain the compression to a limited but varying length range [min_length, max_length] according to the length of the original sentence:

$$min\_length = \min\{original\_length, \alpha\}$$
$$max\_length = \begin{cases} \beta + \sqrt{orig\_length - \beta}, & \text{if } original\_length > \beta \\ original\_length, & \text{otherwise} \end{cases} \quad (3)$$

where
orig_length : the length of original sentence in characters
α, β : fixed lengths in characters

In contrast to a fixed limitation of length, this varying length simulates human behavior in creating compression and avoid the overcompression caused by the density selection schema.

4   Experiments

4.1   Experiment Setup

Our experiments were designed to evaluate the quality of automatic compression. The evaluation corpus is 79 documents from Chinese newswires, and the first sentence of each news article is compressed.

The compression of the first sentences in the Chinese news articles is a comparatively challenging task. Unlike English, Chinese often connects two or more self-complete sentences together without any indicating word or punctuation; this is extremely frequent for the first sentence of news text. The average length of the first sentences in the 79 documents is 61.5 characters, compared to 46.8 characters for the sentences in the body of these news articles.

We compare the compressions generated by four different methods:

- Human [H]: A native Chinese speaker is asked to generate a headline-like compression (must be a complete sentence, not a fragment, and need not preserve original SVO structure) based on the first sentence of each news article. Only word deletion operations are allowed.
- Heuristics [R]: The heuristics-only approach mentioned in section 2.1.
- Heuristics + Word Significance [W]: The approach combines heuristics and word significance. The parameter ω in (1) is set to be 1, which is an upper bound of word's tf-idf value throughout the corpus.
- Heuristics + Word Significance + Length Constraints [L]: Compression is constrained to a limited but varying length, as mentioned in section 2.3. The length parameters α and β in (3) are set roughly to be 10 and 20 characters based on our experience.

4.2   Human Evaluation

Sentence compression is commonly evaluated by human judgment. Following the literature (Knight and Marcu, 2002; Dorr et al, 2003; Clarke and Lapata, 2006; Cohn and Lapata 2007), we asked three native Chinese speakers to rate the grammaticality of compressions using the 1 to 5 scale. We find that all three non-linguist human judges tend to take semantic correctness into consideration when scoring grammaticality.

We also asked these three judges to give a list of keywords from the original sentence before seeing compressions, which they would preserve if asked to create a headline based on the sentence. Instead of a subjective score, the informativeness is evaluated by measuring the keyword coverage of the target compression on a percentage scale. The three judges give different numbers of keywords varying from 3.33 to 6.51 on average over the 79 sentences.

The compression rate is the ratio of the number of Chinese characters in a compressed sentence to that in its original sentence.

The experimental results in Table 2 show that our automatically generated compressions preserve grammaticality, with an average score of about 4 out of 5, because of the use of linguistically-motivated heuristics.

	Compression Rate	Grammaticality (1 ~ 5)	Informativeness (0~100%)
Human	38.5%	4.962	90.7%
Heuristics	54.1%	4.114	64.9%
Heu+Sig	52.8%	3.854	68.8%
Heu+Sig+L	34.3%	3.664	56.1%

Table 2. Mean rating from human evaluation on first sentence compression

Event-based word significance and information density increase the amount of important information by 6% with similar sentence length, but decreases the average grammaticality score by 6.5%. This is because the method using word significance sacrifices grammaticality to reduce the linguistic complexity of the heuristics. Nonetheless, this method does improve grammaticality for 16 of the 79 compressed sentences, typically for those with POS or parsing errors.

The compression rates of the two basic automatic approaches are around 53%, while it is 38.5% for manual compression. This is partially because our heuristics only trim the parse tree

but do not transform the structure of it, while a human may change the grammatical structure, remove more linking words and even abbreviate some words. The length constraint boosts the compression rate of our combined approach by 35% with a loss of 18.5% in informativeness and 5% in grammaticality.

Grammaticali-ty (1~5)	Number of Sentence	Compres-sion Rate	Informa-tiveness (0~100%)
Heuristics > 4.5	45	64.1%	75.9%
Heuristics >= 4	62	54.5%	70.6%
Heu+Sig > 4.5	35	59.8%	81.8%
Heu+Sig >= 4	57	56.7%	75.8%

Table 3. Compressions with good grammar

We further investigate the performance of our automatic system by considering only relatively grammatical compressions, as shown in Table 3. The compressions which receive an average score of more than 4.5 are comparatively readable. The combined approach generates 35 such compressions among a total of 79 sentences, preserving 81.8% important information on average, which is quite satisfying since human-generated compression only achieves 90.7%.

The infomativeness score of human-generated compression also demonstrates the difficulty of this task. We compare our automatically generated event words list with the keywords picked by human judges. 61.8% of human-selected keywords are included in the event words list, thus considered when calculating information significance. This fact demonstrates some success but also potential room for improving keyword selection.

4.3 Some Examples

We illustrate several representative samples of our system output in Table 4. In the first example, all three automatic compressions are acceptable, though different in preserving important information. [W] and [L] concisely contain the WHO, WHAT, WHOM information of the event, while [R] further preserves the WHY and WHEN information.

In the second example, the heuristics-only approach produced a decent compression by keeping only the first self-complete sub-sentence. The weight of word "白宫(White House)" is somewhat overwhelming and resulted in dense compressions in [W] and [L], which are too short to be good. Besides, [W] and [L] in this example

show that not all the prepositional phrases, noun modifiers etc. can be removed in Chinese without affecting grammaticality, though in most cases the removals are safe. This is one of the main reasons for grammar errors in the compression results except POS and parsing errors.

The third example shows how the combined approach overcomes POS errors and how length constraints avoid overcompression. In [R], "纳达尔(Nadal)" is deleted because it is mistakenly tagged as an adverb modifying the action "claim the victory and progress through". Since Nadal is tagged as proper noun somewhere else in the document, its significance makes it survive the compression process. [L] produces a perfect compression with proper length, information and grammar, just as human-made compression. [W] selects a very condensed version of compression but loses some information.

1.
[O] 由于对海域疆界划分各执一词, 为期三日的南北两韩高层军事会谈在今天不欢而散.
Because both sides were immovable on the drawing of maritime borders, a three-day high-level military meeting between North and South Korea broke up in discord today.
[H]两韩高层军事会谈今天不欢而散.
A high-level military meeting between two Koreas broke up in discord today.
[R]由于各执一词, 为期三日的两韩高层会谈在今天不欢而散.
Because both sides were immovable, a three-day high-level meeting between two Koreas broke up in discord today.
[L]两韩高层会谈不欢而散.
A high-level meeting between two Koreas broke up in discord.
[W]两韩高层会谈不欢而散.
A high-level meeting between two Koreas broke up in discord.

2.
[O]白宫今天呼吁尽快派遣核检人员, 以监督北韩关闭其核子反应炉; 白宫是在美国总统布希与南韩总统卢武铉电话交谈过后, 作出此一呼吁.
The White House today called for nuclear inspectors to be sent as soon as possible to monitor North Korea's closure of its nuclear reactors. The White House made this call after US President Bush had telephone conversations with South Korean President Roh Moo-hyun.
[H] 白宫今天呼吁派遣人员监督北韩关闭核反应炉.
The White House today called for inspectors to be sent to monitor North Korea's closure of its nuclear reactors.
[R]白宫今天呼吁派遣人员, 以监督北韩关闭反应炉.
 The White House today called for inspectors to be sent to monitor North Korea's closure of its reactors.
[L]白宫今天呼吁派遣人员, 白宫是, 作出呼吁.
The White House today called for inspectors to be sent. The White House is, made this call.
[W]白宫是, 作出呼吁.
The White House is, made this call.

3.
[O]第四种子乔科维奇退赛, 让原以三比六, 六比一, 四比一领先的第二种子纳达尔获胜过关.
Fourth seed Djokovic withdrew from the game, and allowed second seed Nadal , who was leading 3-6 , 6-1 , 4-1 , to claim the victory and progress through.
[H]乔科维奇退赛让纳达尔获胜过关.
Djokovic withdrew from the game, and allowed Nadal to claim the victory and progress through.

[R]乔科维奇退赛, 让以三比六, 六比一, 四比一领先的第二种子获胜过关.   Djokovic withdrew from the game, and allowed second seed, who was leading 3-6 , 6-1 , 4-1 , to claim the victory and progress through.   [L]乔科维奇退赛让种子纳达尔获胜过关.   Djokovic withdrew from the game, and allowed seed Nadal to claim the victory and progress through.   [W]乔科维奇退赛.   Djokovic withdrew from the game.
4.   [O]中新网 7 月 31 日电陈水扁 30 日质疑岛内司法人员企图介入台地区领导人选举.   Chinanews.com , July 31 On the 30th Chen Shui-bian questioned that members of the judiciary on the island may have tried to get involved in elections for leaders in the Taiwan region.   [H]陈水扁质疑司法人员介入台地区领导人选举.   Chen Shui-bian questioned that members of the judiciary may get involved in elections for leaders in the Taiwan region.   [R]中新网 7 月 31 日电陈水扁 30 日质疑岛内人员企图介入台地区领导人选举.   Chinanews.com , July 31 On the 30th Chen Shui-bian questioned that members on the island may have tried to get involved in elections for leaders in the Taiwan region.   [L]陈水扁 30 日质疑人员企图介入台地区领导人选举.   On the 30th Chen Shui-bian questioned that members may have tried to get involved in elections for leaders in the Taiwan region.   [W]陈水扁 30 日质疑人员企图介入台地区领导人选举.   On the 30th Chen Shui-bian questioned that members may have tried to get involved in elections for leaders in the Taiwan region.
5.   [O]帕蒂尔是印度史上第一位女性总统候选人, 如果她当选, 她将成为印度有史以来的首位女总统.   Patil is India's first woman presidential candidate, if she is elected, she will become India's first woman president in history.   [H] 帕蒂尔是印度史上第一位女性总统候选人.   Patil is India's first woman presidential candidate.   [R]帕蒂尔是印度史上第一位候选人.   Patil is the first candidate in the history of India.   [L]帕蒂尔是候选人, 她将成为印度有史以来的总统.   Patil is the candidate, she will become president of Indian history.   [W]帕蒂尔是候选人.   Patil is the candidate.

Table 4. Compression examples including human and system results, with reference translation (O: Original sentence)

The fourth sample indicates an interesting linguistic phenomenon. The head of the noun phrase "岛内司法人员(members of the judiciary on the island)", "人员(members)" cannot stand alone making a fluent and valid sentence, though all the compressions are grammatically correct. Our human assessors also show a preference of [R] to [L, W] in grammaticality evaluation, taking semantic correctness into consideration as well. This is probably a reason that our combined approach performs worse than heuristic-only approach in grammaticality. The combined approach tends to remove risky constituents, but it is hard for word significance to control this risk

properly in every case. This is another of the main reasons for bad compression.

In the fifth sample, all the automatic compressions are grammatically correct preserving well the heads of subject and object, but are semantically incorrect. This case should be hard to handle by any compression approach.

5 Conclusions and Future Work

In this paper, we propose a novel approach to combine linguistically-motivated heuristics and word significance scoring for Chinese sentence compression. We take advantage of heuristics to preserve grammaticality and not rely on a parallel corpus. We reduce the complexity involved in preparing complicated deterministic rules for constituent deletion, requiring people only to determine potentially removable constituents. Therefore, this approach can be easily extended to languages or domains for which parallel compression corpora are scarce. The word significance scoring is used to control the word deletion process, pursuing a balance between sentence length and information loss. The exploitation of event information improves the mechanical rule-based approach in preserving event-related words and overcomes some POS and parsing errors.

The experimental results prove that this combined approach is competitive with a finely-tuned heuristics-only approach to grammaticality, and includes more important information in the compressions of the same length.

In the future, we plan to apply the compression to Chinese summarization and headline generation tasks. A careful study on keyword selection and word weighting may further improve the performance of the current system. We also consider incorporating language models to produce fluent and natural compression and reduce semantically invalid cases.

Another important future direction lies in creating a parallel compression corpus in Chinese and exploiting statistical and machine learning techniques. We also expect that an abstractive approach involving paraphrasing operations besides word deletion will create more natural compression than an extractive approach.

tions expressed in this material are the authors' and do not necessarily reflect those of the U.S. Government.

References

J. Clarke and M. Lapata, 2006. Models for Sentence Compression: A Comparison across Domains, Training Requirements and Evaluation Measures. In Proceedings of the COLING/ACL 2006, Sydney, Australia, pp. 377-384.

T. Cohn and M. Lapata. 2007. Large Margin Synchronous generation and its application to sentence compression. In the Proceedings of the EMNLP/CoNLL 2007, Pragure, Czech Republic, pp. 73-82.

B. Dorr, D. Zajic and R. Schwartz. 2003. Hedge Trimmer: A Parse-and-Trim Approach to Headline Generation. In the Proceedings of the NAACL/HLT text summarization workshop, Edmonton, Canada, pp. 1-8.

M. Galley and K. McKeown, 2007. Lexicalized Markov Grammars for Sentence Compression. In the Proceedings of NAACL/HLT 2007, Rochester, NY, pp. 180-187.

C. Hori and S. Furui. 2004. Speech Summarization: An Approach through Word Extraction and a Method for Evaluation. IEICE Transactions on Information and Systems, E87-D(1): 15-25.

Z. Huang and M. Harper, 2009. Self-training PCFG Grammars with Latent Annotations Across Languages. In the proceedings of EMNLP 2009, Singapore.

H. Jing. 2000. Sentence Reduction for Automatic Text Summarization. In Proceedings of the 6th ANLP, Seattle, WA, pp. 310-315.

K. Knight and D. Marcu, 2002. Summarization beyond Sentence Extraction: a Probabilistic Approach to Sentence Compression. Artificial Intelligence, 139(1): 91-107.

W. Li, W. Xu, M. Wu, C. Yuan and Q. Lu. 2006. Extractive Summarization using Inter- and Intra-Event Relevance. In the Proceedings of COLING/ACL 2006, Sydney, Australia, pp 369-376.

R. McDonald. 2006. Discriminative Sentence Compression with Soft Syntactic Constraints. In the Proceedings of 11th EACL, Trento, Italy, pp. 297-304.

M. L. Nguyen, A. Shimazu, S. Horiguchi, T. B. Ho and M. Fukushi. 2004. Probabilistic Sentence Reduction using Support Vector Machines. In Proceedings of the 20th COLING, Geneva, Switzerland, pp. 743-749.

K. McKeown, R. Barzilay, S. Blair-Goldensohn, D. Evans, V. Hatzivassiloglou, J. Klavans, A. Nenkova, B. Schiffman and S. Sigelman. 2002. The Columbia Multi-Document Summarizer for DUC 2002. In the Proceedings of the ACL workshop on Document Understanding Conference (DUC) workshop, Philadelphia, PA, pp. 1-8.

K. Parton, K. McKeown, R. Coyne, M. Diab, R. Grishman, D. Hakkani-Tür, M. Harper, H. Ji, W. Ma, A. Meyers, S. Stolbach, A. Sun, G. Tur, W. Xu and S. Yaman. 2009. Who, What, When, Where, Why? Comparing Multiple Approaches to the Cross-Lingual 5W Task. In the Proceedings of ACL-IJCNLP, Singapore.

J. Turner and E. Charniak. 2005. Supervised and Unsupervised Learning for Sentence Compression. In the Proceedings of 43rd ACL, Ann Arbor, MI, pp. 290-297.