Natural Language Watermarking Based on Syntactic Displacement and Morphological Division

Mi-Young Kim, Osmar R. Zaiane, and Randy Goebel
Department of Computing Science, University of Alberta
Edmonton, Canada
{miyoung2, zaiane, goebel}@cs.ualberta.ca

Abstract

This paper explores the method for Korean text watermarking and develops a morpheme and syntax based scheme that a predicate nominal is divided into a nominal and a predicate, and syntactic adverbial can be displaced. Korean, as an agglutinative language, provides a good ground for the morpheme-based natural language watermarking because a word consists of several morphemes, and we can also use the characteristics that Korean permits free word order. Korean word usually consists of a content morpheme and a function morpheme. However, predicate nominal has exceptionally two content morphemes—nominal and predicate--and one function morpheme. So, we can divide a predicate norminal into a nominal and a predicate. In addition, most languages permit displacement of syntactic adverbials within its clause. Combining these two characteristics, we propose a method of language watermarking based on syntactic displacement and morphological division. To make our system more secure, we adopt a sentence weight value and make the weight value carry a watermark bit. Our watermarking method doesn't change the meaning of the most marked sentences, and it also ensures the naturalness of the sentences.

From the experimental results, we show that the rate of unnatural sentences of marked text is reasonable, and the watermarking capacity is better than previous systems. The coverage of marked sentences is also reasonable. Experimental results show that the marked text keeps the same style, and it also has the same information without semantic distortion.

1. Introduction

Text watermarking is an emerging technique in the intersection of natural language processing and the technologies of security. Text watermarking aims at embedding additional information in the text itself with the goals of subliminal communication and hidden information transport, of content and authorship authentication, and finally of enriching the text with metadata [2]. The watermarking techniques have been explored extensively for multimedia documents in the last decade[1]. In contrast, the studies on natural language watermarking have been just starting for the recent several years.

In [3,4,6], the techniques of synonym substitution for watermarking have been addressed and various attack scenarios have been described. In [8], Atallah *et al.* have attempted to use quadratic residues technique to insert a watermark to a given text via synonym substitution. The ambiguity induced on the word precision by the synonym substitution technique has led Topkara *et al.*[10] to syntax-based natural language watermarking. Their technique basically focuses on the syntactic sentence-paraphrasing. In [13,14], they propose morpheme segmentation and syntactic analysis for watermarking. However, they have limitations that the watermarking capacity was very low.

Note that Korean, as an agglutinative language, differs significantly from Indo-European languages such as English with respect that one word consists of several morphemes. For this reason, we believe that Korean and other agglutinative languages provide a good ground for text watermarking based on division of a word by the characteristics of its morphemes. In addition, most languages permit the free order of a syntactic adverbial within its clause boundary. We utilize these two characteristics for text watermarking.

Korean word usually consists of a content morpheme and a function morpheme. However, predicate nominal exceptionally consists of two content morphemes—nominal and predicate—and one function morpheme. Predicate nominal means the predicate that is derived from a nominal. So, this paper proposes text watermarking by dividing a predicate nominal into a nominal and a predicate. We also try to transform a sentence by displacing syntactic adverbials. We embed watermark in original text, creating a ciphertext based on morphological division or syntactic displacement according to the choice of the watermark selector, which preserves the meaning. To make the system more secure, we use a sentence weight and make the weight value carry a watermark bit.

2. Previous Work

M. Atallah *et al.*[7,8] proposed a technique for information hiding in natural language text. Moreover, they established the basic technique for embedding a resilient watermark in text by combining a number of information assurance and security techniques with the advanced methods and resources of natural language processing. A semantically based scheme significantly improves the information-hiding capacity of English text by modifying the granularity of meaning of individual terms/sentences. However, this scheme is suitable only for English, and it was merely conceptual.

A technique of embedding secret data, without changing the meaning of a text a lot is proposed by replacing words in the cover text with synonyms[3,4,6]. However, there is deterioration in documents in which importance is attached to delicate nuance when synonyms have been substituted. There are also cases that wrong words are selected as synonyms among many synonym candidate words. Moreover, the method needs a large synonymy dictionary and a huge collocation database[11].

Some methods proposed the text watermarking for agglutinative languages. H.M.Meral *et al.*[2] proposed morphosyntactic tools for Turkish text watermarking, and O. Takizawa *et al.*[11] suggested the adjustment to new line positions for Japanese text. This method has limitations that the message sender and recipient must share the same secret rule table, and the total rate of embedding is too low.

Topkara *et al.*[10] also proposed syntax-based natural language watermarking using the syntactic sentence-paraphrasing. They insist that the syntactic approach is useful for natural language watermarking without semantic distortion. But, sentence paraphrasing can result in unnaturalness of the sentence.

M. Y. Kim[13] showed watermarking method based on morpheme segmentation and syntactic analysis method was proposed in [14]. However, the embedding capacity was too low. H. Wang et al.[12] proposes watermarking method using Chinese syntactic transformation. They introduce the sentence weight value for each sentence, and embed a watermark bit in the weight value.

We propose a text watermarking method combining morphological division and syntactic displacement, to improve coverage of watermarking. To improve system security, we assign a sentence weight for each sentence, and the watermark bit is hidden in the weight value.

We conclude that morpheme and syntax based combined text watermarking will be effective. It makes the rate of embedding high and the coverage of the system better. It also keeps the meaning and naturalness of a sentence. In the next section, we describe our method in detail.

3. Text Watermarking based on Syntactic Displacement and Morphological Division

3.1. Embedding watermarking

We propose two watermarking schemes for watermarking: one is morphological division, and the other is syntactic displacement. Watermark selector selects sentences that are applicable for watermarking, and chooses the method for watermarking between the two schemes. If both of the two methods can be applied in a sentence, we apply round-robin between two methods to balance their appearances.

A quadratic residue function is adopted for watermarking[12]. An integer q is called a quadratic residue modulo z if it is congruent to a perfect square (mod z); i.e., if there exists an integer y such that: y^2=q(mod z); Otherwise, q is called a quadratic nonresidue (mod z). Z should be a prime number where there are an equal number of residues and nonresidues. The quadratic residue function *qr(key, W)* is given in the following:

qr(key, W)= 1 if it is a quadratic residue (mod *key*)
0 otherwise,

where *key* is a prime number and *W* is a weight value.

We apply the weight value of H. Wang et al. [12] using the characteristics that the positions and lengths of words in a marked sentence are different from the original sentence. Thus, if we make a sentence weight based on the length of each word and its position, the weight value for the whole sentence will change if the sentence is changed. We can utilize the weight value *W* of the original sentence and a new value (either 1 or 0)

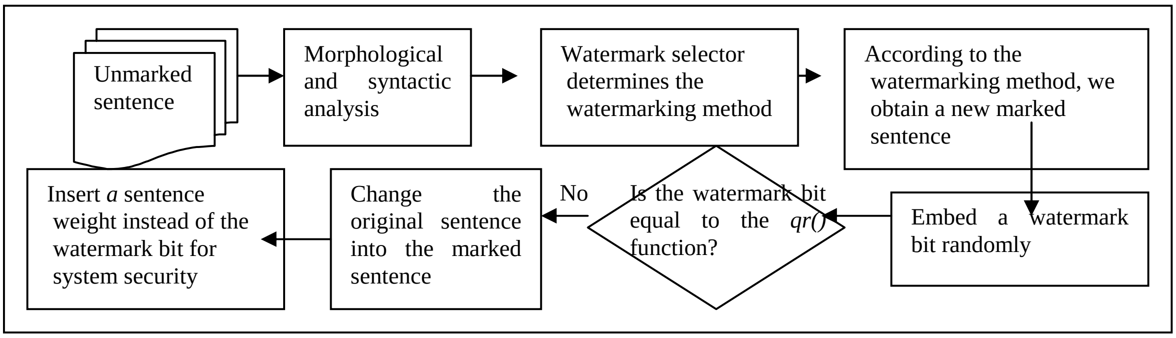

Unmarked sentence
Morphological and syntactic analysis
Watermark selector determines the watermarking method
According to the watermarking method, we obtain a new marked sentence
Embed a watermark bit randomly
Is the watermark bit equal to the qr() function?
No
Change the original sentence into the marked sentence
Insert a sentence weight instead of the watermark bit for system security

Fig. 1. Text watermarking procedure

***Chul-pan-doi-go* : *chul-pan* (common noun) + *doi* (support predicate) + *go* (coordinate ending)**

➔ **Dividing it into two words**

(1) first new word : *chul-pan* (common noun)

(2) second new word : *doi* (support predicate) + *go* (coordinate ending)

➔ **Inserting a functional word to the first word**

(1) *chul-pan* (common noun) + **i(subject case particle)**

(2) *doi* (support predicate) + *go* (coordinate ending)

Original word: **chul-pan-doi-go**

Newly obtained words after morphological division : **chul-pan-i doi-go**

Fig. 2. Example of morphological division

for the sentence will be obtained through the *qr(key, W)* function.

The embedding process of watermarking is as follows.

(1) Morphological and syntactic analyses are carried for all the sentences in the whole text, during which watermarking applicability is determined for each sentence.

(2) The weight value Wj for a sentence Sj is computed as follows.

Sentence weight *W* for a sentence S is defined by:

$$W = \sum_{p=0}^{n} 2^{p} l(p) \ ,$$

where n is the number of words in S, and *l*(p) is the length of (p+1)*th* word. Then, compare its quadratic residue function value with the watermarking bit *Mi*. If equals, no change is taken; otherwise, let watermark selector choose the method for watermarking and obtain a marked sentence until *qr(key,Wj)* equals *Mi*. If all transformations fail, no change is made to *Wj*.

For example, in the sentence of Fig. 3, the *W* value becomes

2^0*5+2^1*1+2^2*3+2^3*3+2^4*3= 91.

(3) Delete the above watermarking bit *Mi* from the watermark and insert *Wj* as used.

(4) Repeat steps (2)-(3) until the whole watermark information is embedded.

Fig. 1 shows the overall procedure of our text watermarking.

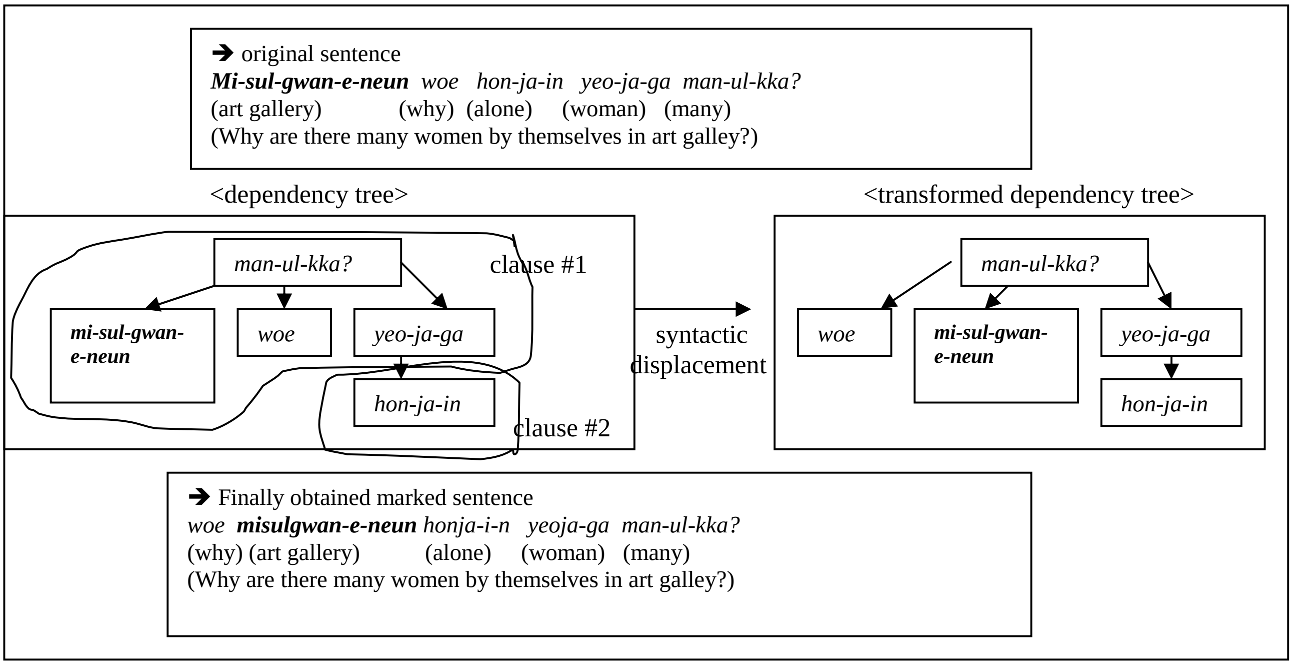
3.2. Watermarking extraction

We input the possibly marked sentence and obtain morphological and syntactic analysis results. The watermark extraction process is similar to embedding process except that in this case we just calculate quadratic residue function value with the prime key and the sentence weights as the input arguments.

Then, we determine the watermarking method that must have been applied.

3.3. Watermarking method : Morphological division

Korean is an agglutinative language that one word consists of several morphemes. Usually one word consists of a content morpheme and a function morpheme. However, some words exist that have more than one content morpheme. So, we try to find the word type that has more than a content morpheme, and divide the word into two new words.



➔ original sentence
Mi-sul-gwan-e-neun *woe hon-ja-in yeo-ja-ga man-ul-kka?*
(art gallery) (why) (alone) (woman) (many)
(Why are there many women by themselves in art galley?)

➔ Finally obtained marked sentence
woe ***misulgwan-e-neun*** *honja-i-n yeoja-ga man-ul-kka?*
(why) (art gallery) (alone) (woman) (many)
(Why are there many women by themselves in art galley?)

Fig. 3. Example of syntactic displacement

In Korean, a predicate nominal has exceptionally two content morphemes – a nominal and a predicate--. We choose the first predicate nominal in a sentence for division. We divide each predicate nominal into two new words, and insert a function morpheme for the first new word that does not have a function morpheme. You can see one example in Fig. 2.

Predicate nominal consists of "first content morpheme(nominal) + second content morpheme (support predicate) + function morpheme".

As shown in Fig. 2, the first new word consists of only a nominal, and the second new word consists of a predicate and a function morpheme – in Fig. 2, coordinate ending is used as a function morpheme. Then, the first new word does not have a function morpheme. So, we insert a relevant function morpheme considering the relation between the nominal and predicate.

We select a relevant function morpheme as following.

(1) If the support predicate indicates active voice (e.g.'*ha', 'siki'*), then the nominal functions as the object of the predicate. So, we insert an object case particle--'*eul*' or '*reul'*.

(2) If the support predicate indicates passive voice(e.g. 'doi'), then the nominal functions as the subject of the predicate. So, we insert a subject case particle --'*i*' or '*ga*'.

In Fig. 2, we can see that a subject case particle is inserted in the first new word because the support predicate('doi') shows a passive voice.

If a sentence can be changed based on morphological division, the watermark selector determines the sentence will be marked using morphological division, and the watermark bit and quadratic residue function value for the sentence is not equal, then we will get a marked sentence based on morphological division.

3.4. Watermarking method : Syntactic displacement

Syntactic dependency parser functions to determine the syntactic relation between words in a sentence. To obtain a syntactic dependency parsing result, we use a Korean syntactic parser of M. Y. Kim *et al.*[5]. Fig. 3 shows an example of a syntactic dependency tree.

Although the Korean language admits relatively free word order, the boundary that a word can move is limited. We can displace a word within the clause that it belongs to. Adverbial, rather than other constituents, can move more freely in a sentence without semantic distortion[9]. Therefore, as a target node for movement, we choose an adverbial constituent from a syntactic tree. We consider the displacement position is among

Table 1 Performances of our system

		Test sentences
The number of sentences		3,000
Avg number of words/sentence		17.95
sentences selected for embedding watermark bit string(%)		82.17%
Naturalness of Marked sentences	Morphological division	99.03%
	Syntactic displacement	92.11%
Unnatural sentences among marked sentences		8.07%
Unsuitable sentences among non-transformed sentences		7.22%

Table 2 Comparison of Performances with other systems

	H. M. Meral[2]	H. Wang[12]	Our system
Coverage of watermarking		11.44%	82.17%
information-hiding capacity	0.81 bit/sentence	0.36 bit/sentence	0.86bit/sentence

Table 3 Comparison of Performances for average edit

	H. M. Meral[2]	Our system
Edit-hit rate (%)	6.1%	7.70%

the positions of the nodes with the same level hierarchy of the adverbial node in a clause.

Two kinds of displacement directions exist – left and right. In this study, we only consider moving to the right position. We choose the first syntactic adverbial that can be moved to the right position in a sentence, and move the adverbial to the right nearest position. In Fig. 3, '*mi-sul-gwan-e-neun*' is an adverbial, and it moves to the right nearest position. Finally, from the modified syntactic tree, we generate a marked sentence as shown in the bottom of Fig. 3.

If a sentence can be changed based on syntactic displacement, the watermark selector determines the sentence will be marked using syntactic displacement, and the watermark bit and quadratic residue function value for the sentence is not equal, then we obtain a marked sentence using adverbial displacement.

4. Experimental Results

We have used 3,000 declarative sentence set in the corpus of Matec99(Morphological Analyzer and Tagger Evaluation Contest in Korean). As shown in Table 1, the average number of words/sentence is 17.95.

We measure subjective rate by human as H.M.Meral *et al.* [2] used. The evaluation method is to let human evaluate the text and show their reactions by editing attempts. The subjects are given marked text and asked to edit them for improved intelligibility and style. This is a blind test because the subjects are not aware that text watermarking has taken place. Three humans have checked the sentences.

We measured the following performances.

1. Coverage of marked sentences
2. Naturalness of marked sentences
3. Human evaluation result
4. Information hiding capacity

We obtained the following results.

1. The coverage about the sentences selected for embedding watermark bit is 82.17% .
2. Naturalness of morphological division is 99.03%.
3. Naturalness of syntactic displacement is 92.11% .
4. The embedding rate(information-hiding capacity) is 0.86 bit/sentence.
5. The percentage of unnatural sentences among marked sentences is 8.07%.
6. The percentage of unnatural sentences among non-transformed sentences is 7.22% .

Table 1 shows the rate of unsuitable sentences among marked sentences and that among untransformed sentences. It is also interesting to note that sentences which have not transformed have also received edit hits at a rate of 8.07%, implying that the edit hits between marked sentences and non-marked sentences are not so different.

The average edit rate is 7.70%, which shows worse result than that of H. M. Meral *et al.*[2] as shown in Table 3. However, the editing rate is very subjective according to the human who edited, and we should also consider that the language and length of sentences are different between two methods.

In Table 2, the information-hiding capacity of our method shows best among the three methods.

We conclude that our natural language watermarking based on morpheme division and syntactic displacement shows reasonable performance without much semantic and stylistic distortion, and this method also shows good coverage. We also show improved watermarking capacity and we use sentence weight value to make our system more secure.

5. Conclusion

We propose natural language watermarking for Korean based on morphological division and syntactic displacement. By using the characteristics that a word in the agglutinative language usually consists of several morphemes, we divide the word that has two content morphemes into two new words, and a new function morpheme is inserted for the first new word that does not have a function morpheme. We also use the characteristics that syntactic adverbial can be displaced within its clause boundary. To improve watermarking capacity, we adopt a sentence weight value and make the weight carry a watermark bit. The experimental results show that the coverage of our method is 82.17%, the average edit rate is 7.70%, and the watermarking capacity is 0.86 bit/sentence, outperforming previous systems.

We conclude that our watermarking method based on morpheme division, syntactic displacement, and the adoption of a sentence weight value is useful in watermarking of Korean text. We will try to apply our method to other agglutinative languages to demonstrate that this method is effective on other languages.

Acknowledgements

(what should I write here?)

References

[1] I.Cox, M. L. Miller, J. A. Bloom, and M. Kaufman, “Digital Watermarking”, 2002

[2] H. M. Meral, B. Sankur, A. S. Oszoy, T. Gungor, E. Sevinc “Natural language watermarking via morphosyntactic alterations”, In Computer Speech and Language 23 pp. 107-125, 2009

[3] M. Topkara, C. M. Taskiran, E. J. Delp, “Natural language watermarking”, *SPIE Conf. On Security, Steganography and Watermarking of Multimedia Contents*, 2005

[4] C. M. Taskiran, M. Topkara, E. J. Delp, “Attacks on linguistic steganography systems using text analysis”, *SPIE Conf. On Security, Steganography and Watermarking of Multimedia Contents*, pp. 313-336, 2006

[5] M. Y. Kim and J. H. Lee, “Two-Phase S-clause Segmentation”, *IEICE Transactions on Information and Systems*, pp. 1724-1736, 2005

[6] U. Topkara, M. Topkara, M. J. Atallah, “The hiding Virtues of Ambiguity: Quantifiably Resilient Watermarking of Natural language Text through Synonym Substitutiions”, In *Proc. Of ACM Multimedia and Security Conference*, 2006

[7] M. Atallah, V. Raskin, C. F. Hempelmann, M. Karahan, R. Sion, K. E. Triezenberg, U. Topkara, “Natural language watermarking and tamperproofing”, *Lecture Notes in Computer Sciences*, 2002

[8] M. J. Atallah, V. Raskin, M. Crogan, C. Hempelmann, F. Kerschbaum, D. Mohamed, S. Naik. “Natural language watermarking: design, analysis, and proof-of-concept implementation”, In *Proc. Of the International Information Hiding Workshop*, 2001

[9] J.I. Kwon, “The study of Korean grammar” Bak-I-Jeong, 1994

[10] M. Topkara, U. Topkara, M. J. Atallah, “Words are not enough: sentence level natural language watermarking”, In *Proc. of 4th ACM International Proceedings of ACM Workshop on Content Protection and Security* (in conjuction with ACM Multimedia), 2006

[11] Osamu Takizawa, Kyoko Makino, Tsutomu Matsumoto, Hiroshi Nakagawa, Ichiro Murase: Method of Hiding Information in Agglutinative Language Documents Using Adjustment to New Line Positions. *KES* (3) pp. 1039-1048, 2005

[12] H.Wang, X. Sun, Yu. Liu and Yo. Liu, “Natural Language Watermarking Using Chinese Syntatic Transformations”, *Information Technology Journal* 7 (6): 904-910, 2008

[13] M. Y. Kim, “Natural language watermarking by morpheme segmentation”, *Proc. of the Asian Conference on Intelligent Information and Database Systems*, pp.144-149, 2009

[14] M. Y. Kim “Text watermarking by syntactic analysis”, *Proc. of WSEAS International Conferences on Computers*, pp.904-909, 2008