ABSTRACT—This letter addresses a structural disambiguation method for Korean adverbs based on the correlative relation constraints between adverbs and modifiees, and the morphological context information of sentences. Using the proposed method, we improved the dependency parsing accuracy of adverbs from 79.2 to 89%. The experimental result shows that the proposed method is especially expert in parsing adverbs which can modify multiple word classes or have a long distance dependency relation to their modifiees.

Keywords—Adverb parsing, structural disambiguation, correlative relation, morphological context.

I. Introduction

Adverbs have various syntactic roles; they can modify verbs, adjectives, and other adverbs in addition to other more specialized functions. However, dealing with adverbs has been given less attention in natural language processing research than other word classes though some linguists have done significant work on adverbs. Since S. Colon and others [1] proposed the adverbial lexicon to enable computers to handle adverbs, there have been several studies on adverbs [2], [3]. But these studies concentrated on the generation of adverbs in machine translation rather than parsing.

In Korean, because of its free-word-order property, the position of adverbs in a sentence is relatively free. This freedom with regard to word order combined with the large number of syntactic roles of adverbs makes the structural analysis of Korean adverbs more complicated. For this reason, errors in parsing adverbs often cause parsing failure. In general, Korean parsers have used the information regarding the word class and the position of an adverb and its modifiee in adverb parsing; however, this approach cannot cover various uses of adverbs. Recently, a few studies have dealt seriously with the parsing of Korean adverbs [4], [5]. In [4], S.E. Shin and others propose an adverb parsing method using a modifier dictionary. A study by S.Y. Hwang and others [5] handles the noun-modifying adverbs. However, these previous works are unable to handle adverbs with a long-distance dependency relation [6] to modifiees or adverbs which can modify multiple word classes.

In this letter, we propose a structural disambiguation method for Korean adverbs based on the correlative relation between adverbs and modifiees, and the morphological context of sentences. The role of an adverb in a sentence is to modify and restrict one or more other sentence component(s), so it has a close correlative relation to its modifiee(s) in comparison with other parts-of-speech. For some adverbs with a weak correlative relation, we also use morphological context information. We extracted this analysis information from Electronics and Telecommunications Research Institute (ETRI) tree-tagged corpus. Section II explains the correlative relation between adverbs and modifiees, and the morphological context information. We also introduce a new model for parsing adverbs in section II. Section III verifies the validity of the model compared to a baseline model by showing experimental results, and section IV concludes this work.

1) Adverbs generally modify verbs, adjectives, clauses, and sentences, but not nouns; however, some linguists argue that some adverbs show that they can modify nouns. For more details, see http://www.cis.upenn.edu/~xtag/tech-report/node167.html. In this letter, we accept the idea of noun-modifying adverbs. In Korean there are many more noun-modifying adverbs than in English.
II. Adverb Parsing Model

1. Correlative Relation Constraints Based Analysis

The Korean adverb gyeolko ‘never’ modifies a verb. Example sentences using gyeolko are given in (i) and (ii). The adverbs and their modifiees are indicated with a bold font in example sentences. In (i), the adverb modifies the closest verb bad ‘receive’. But in (ii), it modifies the verb dongjeongha ‘commiserate’ not the adjacent verb ul ‘cry’.

(i) 그는 결과 결과에 영향을
geneu noun gyeolko gyeolkwa-e yeonghyang-eul
he-Nom2) never result-by influence-Acc3)
받지 않았다고 결심했다.
bad-jian-ket-dago geolsim-hayeot-tta.
receive-Neg-Will-ConE decide-Past-End.
(He decided not to be influenced by the result.)

(ii) 그는 결과 온고 있는 소녀를
geneu noun gyeolko ul-goit-neun sonyeo-lead
he-Nom never cry-PrenE5) girl-Acc
동정하지 않았다.
dongjeonghaji-an-aht-da.
commiserate-Neg-Past-End.
(He never commiserated with the crying girl.)

The following sentences illustrate the usage of Korean adverb gajang ‘most’, which can modify a noun, an adjective, an adverb, and a verb. In (iii), gajang modifies the immediate next noun dwi-e ‘back’. In (iv), however, it modifies the adjective chakan ‘kind’ not the noun maeum-i ‘mind’.

(iii) 그는 가장 뒤에 있는
geneu noun gajang dwi-e it-neun
he-Nom most back-at exist-PrenE
책상에 있었다.
chaegssang-e unji-at-tta.
table-Acc sit-Past-End.
(He sat at the rearmost desk.)

(iv) 그는 반에서 가장 마음이
geneu noun ban-es eo gajang maeum-i
He-Nom class-in most mind-Nom
착한 사람이다.
Chakan saram-i-da.
kind person-be-End.
(He is the kindest person in his class.)

These examples show that the modifiee of an adverb is not determined only by the word class and the position of an adverb and its modifiee which are generally used in parsing Korean adverbs. We can get some clues from these examples. The adverb gyeolko ‘never’ has a negative meaning; so it modifies only a sentence component with a negative meaning. The adverb gajang ‘most’ belongs to an adverb of degree; it modifies only the component of which meaning has a stage of degree. Using these restrictions on correlative relation, we can resolve the disambiguation of adverb parsing in (i) to (iv). We call these a part-of-speech correlative relation constraint and a semantic correlative relation constraint, respectively.

We also use lexical constraints and function-word constraints in disambiguation. Lexical constraint is a kind of a fixed expression. Sentences (v) to (vi) show the function-word constraints. In (v), the adverb manyak ‘if’ correlates with the connective ending damyeon ‘if’ combined to verb teona ‘leave’; in (vi), the adverb ama ‘as if’ correlates with the auxiliary verb eulgeosi ‘will’ combined to verb gajyeoga ‘take’. All adverbs in (v) to (vi) modify eojeols satisfying these function-word correlative relation constraints of each adverb.

(v) 만약 그녀가 떠났다면
Manyak geunyeo-ga teona-ss-damyeon
if she-Nom leave-Past-if
그는 그녀를 만나지 못할 것이다.
geun noun geunyeo-lead manna-jimot-halgeosi-tta.
he-Nom she-Acc meet-Neg-will-End.
(If she left, he would not meet her.)

(vi) 비가 왔으니 그는 아마
Bi-ga o-ass-euni geun noun ama
rain-Nom come-Past-because he-Nom maybe
우산을 가져갔을 것이다.
u san-eul gajyeoga-ass-eulgeosi-da.
umbrella-Acc take-Past-will-End.
(Because it rained, maybe he would have taken an umbrella.)

An eojeol is a surface level form consisting of more than one combined morpheme. We use many complex morphemes that are generally not accepted as morphemes in traditional Korean grammar. They are a combined form with several morphemes; they become one syntactic or semantic unit as a combined form. For example, eulgeosi ‘will’ in (vi) is a complex auxiliary verb composed of a prenominal ending eul, a dependent-noun geos ‘thing’, and a copular verb i ‘be’. Therefore, the Korean sentence component called eojeol in this letter is a bigger unit than the general Korean eojeol. We consider the adverb ama in (vi) as modifying the complex eojeol gajyeoga-ass-eulgeosi-da.
2. Morphological Context Based Analysis

Some adverbs such as conjunctional adverbs and modal adverbs have a weak correlative relation to modifiees, so sometimes they cannot be parsed correctly by using correlative relation constraints. To solve this problem, we carry out a low-level sentence understanding using morphological context information, which is sequence information concerning a morpheme which is located before or behind an adverb and a modifiee. The followings are examples of morphological context information for the conjunctional adverb ‘geuleona’ ‘but’ and the modal adverb, ‘teuki’ ‘especially’.

(a) V1+prenominal-ending geuleona V2+ prenominal-ending
→ V2 is the modifiee

(b) NP1, teuki NP2 (NP1, especially NP2)
→ NP2 is the modifiee

Sentences (vii) and (viii) can be correctly parsed by the morphological context information of (a) and (b).

(vii) 그는 비록 마른 그러나
geu-neum bijjeok-na-leun geuleona
healthy-PrenE be_person-End.
(He is a thin but healthy person.)

(viii) 그는 물고기, 특히 연어를
geu-neun mulgogi, teuki yeon-eo-lul
healthy-PrenE fish especially salmon-Acc
like-End.
(He likes fish, especially salmon.)

3. Adverb Parsing Model

Some adverbs have positional constraint on modifying modifiees. Considering this as a positional correlative constraint to composing a correlative relation between an adverb and a modifiee, we resolve ambiguities using the positional constraint when there are more than two candidates satisfying the adverb correlative relation constraints described in section II.

Let the subscripts \( p, s, g, c \) and \( l \) denote the part-of-speech correlative relation constraint, the lexical or semantic correlative relation constraint, the function-word correlative relation constraint, the morphological context information, and the positional constraint of an adverb \( a \), respectively. And also, let \( f_p, f_s, f_g, f_c, \) and \( f_l \) respectively describe the satisfaction function of each constraint condition \( p, s, g, c, \) and \( l \) for \( h_n \) an element of head candidate set \( H \). Each satisfaction function takes 1 if a given constraint is satisfied or there is no constraint. In other cases, the function takes 0. The candidate with the highest selection score becomes the head of the adverb \( a \). The following are the equations for calculating the selection score.

\[
\text{Score}(a, h_i) = c_1 f_{p}(a, h_i) + c_2 f_{s}(a, h_i) + c_3 P(R_{a\_pogel}),
\]

\[
f_p(a, h_i) = f_p(a, h_i) \times f_s(a, h_i) + f_g(a, h_i) \times f_c(a, h_i),
\]

where \( h_i \in H, 1 \leq i \leq |H| \) and \( c_1, c_2, c_3 \) are weight factors.

In the equation, \( P(R_{a\_pogel}) \) describes the probability that adverb \( a \) has a dependency relation \( R_{a\_pogel} \) with an eojeol that satisfies all the given constraints \( p, s, g, c, \) and \( l \). Its value is calculated by (3) from the tree-tagged corpus, where \( freq_a \) denotes the frequency of adverb \( a \) in the corpus and \( freq_{R_{a\_pogel}} \) is the frequency of organizing the dependency relation \( R_{a\_pogel} \) in respect to adverb \( a \).

\[
P(R_{a\_pogel}) = \frac{freq_{R_{a\_pogel}}}{freq_a}.
\]

### III. Experimental Results

We experimented on the ten most frequently occurring adverbs in the Korean Information Based System (KIBS) corpus. Two hundred sentences were part of the experiment for each adverb. The average length of sentences is 17.65 eojeols. The restriction features for each adverb were extracted from the ETRI tree-tagged corpus. In the experiment, the parameters \( c_1 \) and \( c_2 \) were 1 and \( c_3 \) was 0.5. The result is shown in Table 1.

### Table 1. Parsing accuracy of high frequency adverbs

<table>
<thead>
<tr>
<th>Adverb lexeme</th>
<th>Baseline model</th>
<th>Proposed model</th>
</tr>
</thead>
<tbody>
<tr>
<td>geuleona ‘but’</td>
<td>93% (186)</td>
<td>97% (194)</td>
</tr>
<tr>
<td>geuligo ‘and’</td>
<td>28% (56)</td>
<td>49% (98)</td>
</tr>
<tr>
<td>gajang ‘most’</td>
<td>96% (192)</td>
<td>98% (196)</td>
</tr>
<tr>
<td>deo ‘more’</td>
<td>91% (182)</td>
<td>99% (198)</td>
</tr>
<tr>
<td>titalaseo ‘therefore’</td>
<td>75% (150)</td>
<td>84% (168)</td>
</tr>
<tr>
<td>bado ‘just’</td>
<td>19% (38)</td>
<td>68% (136)</td>
</tr>
<tr>
<td>dasi ‘more’</td>
<td>94% (188)</td>
<td>95% (190)</td>
</tr>
<tr>
<td>gachi ‘together’</td>
<td>100% (200)</td>
<td>100% (200)</td>
</tr>
<tr>
<td>imi ‘already’</td>
<td>96% (192)</td>
<td>100% (200)</td>
</tr>
<tr>
<td>ial ‘well’</td>
<td>100% (200)</td>
<td>100% (200)</td>
</tr>
<tr>
<td></td>
<td>79.2% (1584)</td>
<td>89% (178)</td>
</tr>
</tbody>
</table>
The baseline model, which was used in comparison, parses adverbs based on the part-of-speech and the position of the adverb and modifiee which is the information generally used in parsing Korean adverbs. The proposed model enhanced the gross precision of adverb parsing by 9.8%. The proposed method is especially useful in parsing adverbs which can modify multiple word classes or have a long distance dependency relation to their modifiees.

IV. Conclusion

To improve the parsing accuracy of adverbs, we have proposed a structural disambiguation method for Korean adverbs based on the correlative relation between adverbs and modifiees, and the morphological context of sentences. This method improved the dependency parsing accuracy from 79.2 to 89%. We found that many parsing failed adverbs in the experiment could be corrected by understanding sentence meaning more deeply. In future work, we will utilize the extended context information for deeper sentence understanding.

References