

\*

\*\*

- 1.
2.
  - 2.1
  - 2.2
3.
  - 3.1
  - 3.2
  - 3.3
  - 3.4
4.
  - 4.1
  - 4.2
  - 4.3 , ,
- 5.

가 . 가

가 . 가 .

가 . 가

가 ,

가 . 가

(natural language processing)

가 .

1.

가

(artificial intelligence)

가 가

가

가

가 (expert system),

가

(information retrieval system)

가

( ) ,

( )

가

가 ,

가 .1)

(1)

ㄱ.

[ ] [ ]

[ ] [ ]

ㄴ.

[ ] [ ]

( )

(1ㄱ)

가

(1ㄴ)

2),

(1ㄴ) ‘ ’

(speech

recognition) , ‘ ’

(speech syn- thesis)

.3)

(2)

ㄱ.

1)

가 , 가

(1994)

2)

3)

가

가, 가

ㄴ. / + / , / + / ,  
 가 +L/ + /  
 ㄷ. . / /  
 : [ [ 가 ] ]  
 : [ : ] [ : ] [ : ]

(2ㄴ)

(2ㄷ)

, , , , ,  
 . , ,

가 . ,

(2ㄴ) ‘ ’ ‘가 +L / ’ ,  
 +L/ + / ’

. , ,  
 . , 가 ,

(target language)

(source language)

.4)

## 2.

### 2.1

가 .

< >(1988)

, ‘ , ‘ ,  
 . ,  
 .5)

(3) ,

4)

‘ ’ ‘at, on, in’ .

[ ], [ ] )

5)

(1999)

(3)

ㄱ. ,가 , , , , , , , , , , ,  
ㄴ. , , , , , , , , , , , , ,  
ㄷ. , , , , , , , , , , , , ,

(3ㄱ) , (3ㄴ) , (3ㄷ)

. 가 .  
. 가 , , , , , .  
. , , , , .  
가 6) 가 .  
가 .  
. 가 .  
. , .  
. , .  
. , .  
. 가 .  
>(1988) (1996) . <

2.2

가 . 7)  
가 (前處理) 가 .

(4) 1


(4) 가 (delimiter) . 8)

---

6) (character) 가 .  
가 가 .  
(1996) 7 .

7) (1994) , .  
(1998) .

(4) 가 ,

(5) .

(5) 2


(5) ,

(6)

- (6)
- ㄱ. : , ,
  - ㄴ. :

(6~1) " |, |12,340 |, | . |  
 , , 9) .

(matching)

가

가

가

(7)

- ㄱ. We are linguists.
- ㄴ. .
- ㄷ. linguists
- ㄹ. , ,

(7)

[linguists.] |

가 . (7ㄷ, ㄹ)

'linguists' ' , , '가

가 . [linguists.]

8)

9)

||

가 가 , | . | ‘  
가 가 ,  
가

.10)

(5)

3.

, ( , , 가 , , , , , , , ), ( , , , ) 가 ,

3.1

가 , , 가 가 가

- (8) 가. 10:20
- 가'. | | 10:20 |
- 나. 10: 20
- 나'. | | 10: | 20 |
- 다. 10 : 20
- 다'. | | 10 | : | 20 |

(8) 가 . (8가) 2 , (8나) 3 , (8다)

4

, .11) < >  
가 . (9)

- (9) 가. (0) / (0), (0) / (0)
- 나. (0) / (0), (0) / (0)

---

10) 가  
11) (1998)

다. (0) / (0)  
 라. (0) / (0) / (0)  
 마. (0) / (0)

, 가 ,  
 , 가 .

(9-1) 가. | || | | || || |  
 나. [ ] [ ]

(9-1 가) . (9-1 가) ‘ 가 ’  
 (9-1 나)

### 3.2

, , 가 , , , , ,

11) (10, 11, 12) 가 (12) 가 (10, 11)  
 (10) 가 (11) 가

(10) 가. 가 ?  
 나. , ! [ ]  
 다. , , . 5, 6 1, 2, 3, 4 [ ]  
 라. , ; , [ ]

(11) 가. . [ ]  
 나. : , , , [ ]  
 다. / /125

(12) 가. (?) . [ ]  
 나. • , • 가 . [가 ]  
 다. - - - -

(10, 11) (10-1) |  
 ?| | | ?| , (11-1) | . | | |

(12-1) ‘ ( ’ . | (?)| |

| (?) | (12ㄴ) 가  
, (12ㄷ)

(13) “ ..... ” 가 . [ ]

(14) ㄱ. 1919. 3. 1.  
ㄴ. 1. ㄱ. 가.  
ㄷ. . 1987. 3. 5.

(15) ㄱ. 3.14158  
ㄴ. 14,314  
ㄷ. 8 • 15  
ㄹ. 10:20 3:16 65:60  
ㅁ. 3/4 3/20 1996/10/9  
ㅂ. 02-3290-2499 690930-1111111

(13) 가 (|.....|)  
가 (14) 가  
, 가 (ㄱ, ㄴ, ㄷ, 가,  
, ) 가 .12) (15) 가 가

### 3.3

, , 가 가  
가 , , 가  
(( ), { }, [ ]), (--)  
.13) 가

.14)

---

12) (14ㄷ) | .| < > (1996: 36)

13) ‘\_’ , ‘\_’ 가 ‘\_’  
(n dash), (M dash), M dash two M dash 가



(16)

- (16) ㄱ. “ ”
- ㄴ. “ ” 가
- ㄷ. “ , ”
- (17) ㄱ. |" | , |" | , |" |
- ㄴ. | :| , | :| , | :|

(17) . (17ㄱ) ,  
 (17ㄴ) ‘ , ’ 가

- (18) ㄱ. ‘ ’ ‘ ’
- ㄴ. “ ! ” ‘ ’
- (19) ㄱ. | ‘ ’ | .|
- ㄴ. | ‘ ’ | :|

(18) , (19) . (18  
 ㄱ) ‘ ’ 가

19ㄱ) ‘ , ’ ( 19ㄴ) . (19ㄱ) | ‘ ’ |

가  
 , (16ㄴ,  
 ㄷ)

(20, 21)

- (20) ㄱ. (1) (ㄱ) ( )
- ㄴ. ( )
- (21) ㄱ. 3 • 1 (1919)
- ㄴ. (coffee)
- ㄷ. ( )
- ㄹ. (無情) (6 • 25 )
- ㅁ. ( )

(20ㄱ) 가 , 가

(20ㄴ) ) .|

(21)

< >

A. 가

A1. 가 가

| (1919)| | |(1919)| | |1919|

A2. 가

| (coffee) | | (coffee) | | |(coffee) | | |coffee |  
( )

B.

| ( | | ) | | ( | | )| | || | | |  
| ( | | ) | | ( | | )| | | | | |

A2 B

B (21ㄴ, ㄹ)

가

(22) ㄱ. [年歲]가

ㄴ. [ (斷乎) ]

(22)

(21)

/

(21ㄱ) ‘1919’

‘3·1 ’

‘3·1 ’

가‘

’ ‘3·1 ’

가

(23) ㄱ.

ㄴ. | -- | | | | , | -- | | | |

(23ㄴ)

3.4

```

.15)
< >
if: 가 가?
{ < > 가 , < > 가 . }
< >
if: 가 가?
{ 가 , < > 가 . }
else if: 가 , -1
-2 가 (ㄱ, ㄴ, ㄷ, 가, , ) 가?
{ 가 , < > 가 . }
else:
{ , < > 가 . }
< >
if: 가 ( , ) 가?
{ , < > 가 . }
else if: 가 가?
{ 16 < > , < > 가 .
}
else if: 가 가?
{ , . < > 가
. }

```

4.

3

가

가

15)

---

if, else if, else , if 가 , else if  
if , else else if if 가 , else if  
{ }

4.1

가

, , 가 . 가

(24,

25) , (26, 27)

(24)

(25)

(26) 가. 1919. 3. 1.

나. 1. 가. 가.

다. 3.14158

르. . 1987. 3. 5.

르'. G. W. F. Hegel .

(27) 가. “ , ”

나. “ 가 .”

(26-가, 나, 다)

(26르)

. (26르) | .|

.( 1996: 36) (26르')

가

(25) | ., | .|

(28) 가. . . [ ] [ ]

나. 1. 가. 가. [ ] [ ]

다. . [ 1] [ ]

르. G. W. F. [ 2] [ ]

(28)

( 28-가)

( 28나, 다, 르)

가

(28-가)

( , )

( , ,

)

(28나, 다, 르)



(32)

가

(32)

가

(31, 32)

< , , >  
 -1 가  
 -2 가  
 \*\* 가 -1

< >  
 if: 가 가?  
 { < , , > 가 . }  
 else:  
 { < > 가 . }  
 < >  
 < 2>  
 if: ( , , )가 가?  
 { < > 가 , < > 가 , < > 가 . }  
 < >  
 if: , 가  
 (¬, ∩, ∪, 가, , , A, B, C, a, b, c ) 가?  
 { 가 , < 2> 가 . }  
 else if: 가?  
 { 가 , < 2> 가 . }  
 else:  
 { , < 2> 가 . }  
 < >  
 if: 가 가?  
 { 가 , < 2> 가 . }  
 else:  
 { , < 2> 가 . }  
 < >  
 if: 가 가?  
 { 가 , < 2> 가 . }  
 else:  
 { , < 2> 가 . }

4.2

가 , 가 가

- (33) ㄱ. , ?  
 ㄴ. ,  
 ㄷ. ,

(33ㄱ, ㄴ) 가 , (33ㄷ) 가 . (33ㄷ)  
 가

(33ㄱ, ㄴ) 가 가 (34ㄴ)  
 가

- (34) ㄱ. ,  
 ㄴ. [N] [N] [ ] [ ] [NP] [ ]

‘ 가 , 가 ’  
 가 .16)

- (35) ㄱ. ,  
 ㄴ. ,

(35) ‘ ’ ‘ ’가 , ‘ ’ ‘ ’  
 가 가 가 ‘ ’ (35ㄴ) ‘ ’ ‘ ’  
 (35ㄱ) 가 가

(35ㄴ) ‘N1 N2 N3’ ‘[N1 N2] N3’ ‘N1 [N2 N3]’  
 가 가 ‘N1 [N2 N3]’ N2 N3 N3가 N1  
 N3가 ‘ 가 가  
 가

---

16) 가 가

(36) ㄱ. , , 가 , .  
ㄴ. , , 가 , .

(36) 가 , ( 2 )가 2 .

< >  
if: 가 , 가 가?  
{ . }  
else if: 가 , 가 가?  
{ ,가 가 . }  
else if: ( 2 )가 , 가 2 가?  
{ . }

4.3 , ,

가 가 가가

(37) ㄱ. .  
ㄴ. .  
ㄷ. 가 , .  
(38) ㄱ. 가 ?  
ㄴ. 가 가 ?  
(39) ㄱ. , !  
ㄴ. !  
ㄷ. !  
ㄹ. !

(37, 38, 39) , , (38)  
가 . , (38)

(40) ㄱ. 가 .  
ㄴ. ?  
ㄷ. , !  
(40) 가 가





< >

(1999). < >. :

(1996). < >.

(1994). < >.

(2001). < >.

(1988). < >.

(1999). < >. :

, (1998). “ .” < > 37.

(1998). “ .” < > 2.

(2001). “ - .” < > 1.

(1998). “ .” < > 37.

<Abstract>

## **Natural Language Processing and Punctuation Marks**

Kyung-Ho Yi  
Korea University

We very often use punctuation marks in processing of communication with others in written language. And up until now punctuation marks are interpreted only by human beings. But, metaphorically speaking, computer, a new interpreter, became the new user of punctuation marks. The advent of the new age was possible by virtue of natural language processing. In order to help understanding this situation, I explain how punctuation marks are processed in natural language processing.

In Chapter 1, I briefly examine natural language processing in general. In Chapter 2, I explain the information encoded in punctuation marks and I think of the role and the place of punctuation marks processor. Further, I argue for the stages of punctuation mark processing classified into pre-morphological analysis stage and in-syntactic analysis stage. This classification reflects the levels of information that punctuation marks have.

In Chapter 3, I deal with the processing of punctuation marks for the morphological analysis. This process means the operation of eliminating obstacles from morphological analysis. In this chapter, punctuation marks are classified into space character, single marks and pair marks. Single marks function by themselves and pair marks so as pair. Pair marks have scope of operation and the scope often appear in multi word segments, which cause difficulties in processing.

In Chapter 4, I deal with the processing of punctuation marks for the syntactic analysis. In many topics, I focus on the determination of one sentence, comma and the relationship with punctuation marks and types of sentence. First, the determination of one sentence means the process that input strings are divided into one sentence as a unit of syntactic analysis. This process deals with dot, question mark and exclamation mark. And quotations increase difficulties of processing. Second, commas have much information of syntactic structures, such as sentence components and modification. I focus on availabilities of the information in terms of language processing. Finally, in spite of subsidiary conditions, we use punctuation marks in order to determine sentence types. That is to say, if final endings have ambiguity of sentence type indication, we use punctuation marks as a determiner of sentence types.