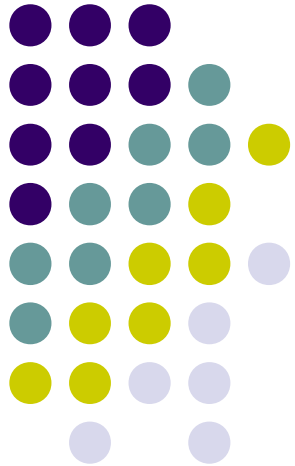
Korean Parsing in an Extended Categorial Grammar



RASLAN 2011

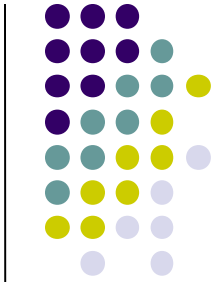
2nd-4th Dec 2011

By Juyeon Kang

2nd of Dec 2011

FI Masaryk University, Brno, Czech Republic

Outline



- Korean Language
- Linguistic characteristics of Korean

1. Motivation
2. Theoretical framework : Applicative Cognitive Grammar
   - CG Formalisms
   - Combinatory Logic
   - ACCG
3. ACCG parser
4. Conclusion

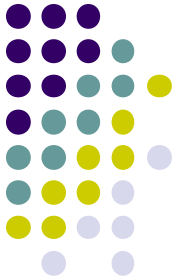
About Hangul

- **Hangul** is the name of the Korean language.
- It was crearted by king Sejong, the 4th ruler of the Chosun Dynasty in 1443.

(1) Basic Consonants and Vowels(24 letters) =Consonants(14 letters) + Vowels(10 letters)

Consonants			
Letter	Name [Pronunciation]	Letter	Name [Pronunciation]
ㄱ	giyeok [kijʌk]	ㅏ	a [a]
ㄴ	nieun [niɯn]	ㅑ	ya [ja]
ㄷ	digeut [tigɯt]	ㅓ	eo [ʌ]
ㄹ	rieul [riɯl]	ㅕ	yeo [jʌ]
ㅁ	mieum [miɯm]	ㅗ	o [o]
ㅂ	bieup [piɯp]	ㅛ	yo [jo]
ㅅ	siot [ʃiot]	ㅜ	u [u]
ㅇ	ieung [iɯŋ]	ㅠ	yu [ju]
ㅈ	jieut [ʧiɯt]	ㅡ	eu [ɯ]
ㅊ	chieut [ʧʰiɯt]	ㅣ	i [i]
ㅋ	kieuk [kʰiɯk]		
ㅌ	tieut [tʰiɯt]		
ㅍ	pieup [pʰiɯp]		
ㅎ	hieut [hiɯt]		

Linguistic characteristics of Korean



1) Typical SOV language

2) Variable word order

ex1 : **eoje** **Eugene-i** **show-reul** **bo-ass-da**

yesterday Eugene-NOM show-ACC watch-PS-DC

Eugene watched a show yesterday.

ex2 : **eoje** **show-reul** **Eugene-i** **bo-ass-da**

yesterday show-ACC Eugene-NOM watch-PS-DC

Eugene watched a show yesterday.

3) Word order typology :

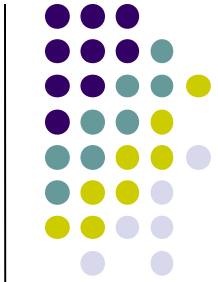
Dem+Num (+CL) +ADJ+N, GenN+HeadN, Rel Cl+N, Verb stem+Aux, N+Case marker

ex : **jeo** **se gae-uy** **jakeun** **inhyeong**

Dem three-CL-GEN little doll

those three little dolls

Linguistic characteristics of Korean (Cont.)



5) Double nominative/accusative construction, double topic construction

ex. Double topic :

Minju-neun umma-neun geurip-da

Minju-TOP mother-TOP miss-DC

Minju missed mother (but not father)

ex. Double accusative:

Minju-ga Inho-reul son-eul jap-ass-da

Minju-NOM Inho-ACC hand-ACC hold-PS-DC

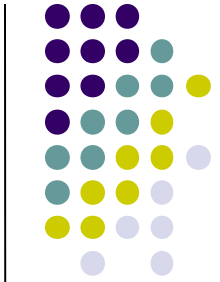
Minju held Inho's hand.

6) Verbs consists of a lexical stem and suffixs

ex : **joah-ass-da (*liked*)**

like-Past-Declarative ending

1. Motivations

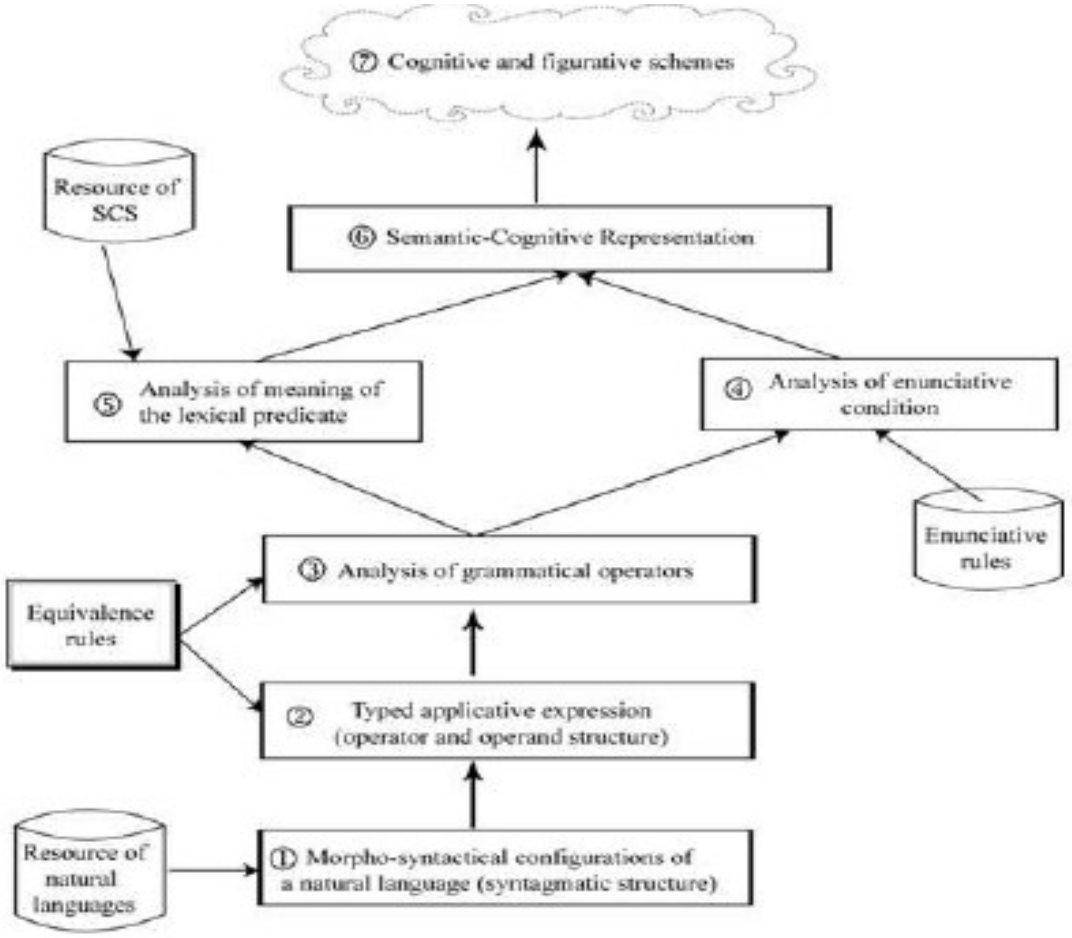


- **Formal analysis of different complex syntactic problems for an efficient parsing**
- **Argument the anti-anti relativism hypothesis (epistemological point of view)**
- **Develop a categorial analysis system adequate to the processing of complex syntactic problem**

2. Theoretical framework :

Cognitive and Applicative Grammar (CAG)

- **Analogue to a compiling program with 7 interrelated levels of representations [Desclés, 2005]**

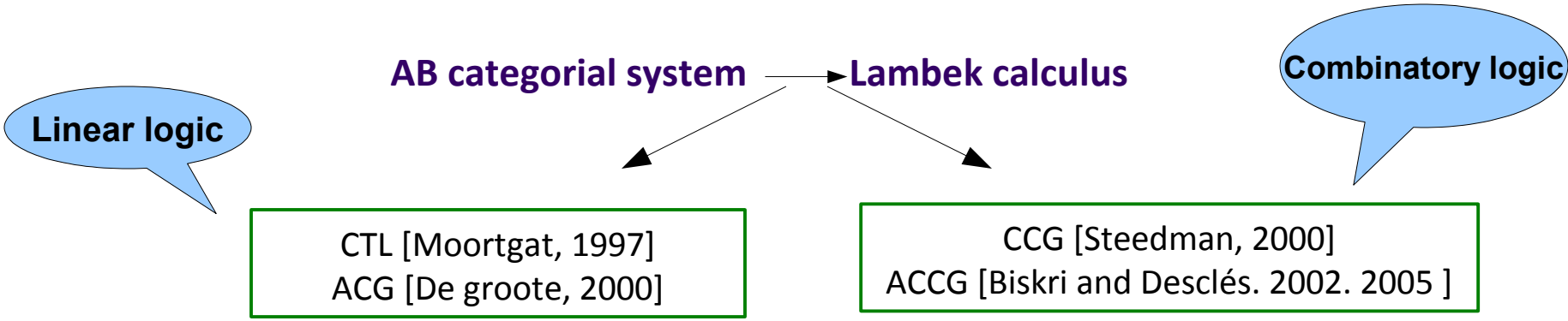


CG formalisms

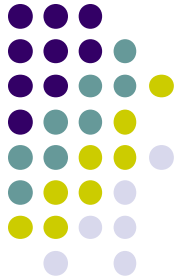
- **Systems of types analogue to Church's functional types**
- **Instance of types are linguistic units analyzed as operators and operands**
- **Calculus on types → application of operators to operands**
- **Basic functional types :**

  S for sentence, **N** for noun and **N*** for the noun phrase

- **<u>Two axes of CG extension</u>**



Combinatory Logic



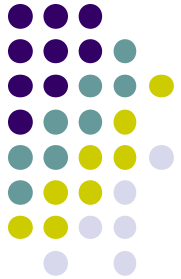
Logic of operators with abstarct operators, « combinators »

- **Operators with assigned types can be composed by different ways [Curry and Feys, 1958]**
- **Combinators are introduced and eliminated by β-reduction rules in Gentzen's style [Fitch, 1974]**

$$\mathbf{B}xyz \geq_{\beta} x(yz)$$

$$\frac{\mathbf{B}fxy}{f(xy)}\,[e\text{-}\mathbf{B}] \qquad \frac{f(xy)}{\mathbf{B}fxy}\,[i\text{-}\mathbf{B}]$$

Applicative Combinatory Categorial Grammar



- **Extension of Combinatory Categorial Grammar [Biskri and Desclès, 2005]**
- **Canonical associations between categorial rules and combinators without binding variables**

Functional composition

$[X/Y : u_1]$-$[Y/Z : u_2]$ ----------**>B** $[X/Z : (\mathbf{B}\, u_1\, u_2)]$

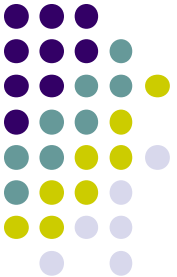
$[Y/Z : u_1]$-$[X\backslash Y : u_2]$ ----------**<Bx** $[X/Z : (\mathbf{B}\, u_2\, u_1)]$

- **Combinators controls : two types of meta-rules**

  combinatory meta-rules and contextual meta-rules
- **Use of schema type (e.g. X/X for the operation of determination)**
- **Module integrated in the CAG**

Meta-rule system

- **Contextual meta-rules (*POS Tagging ambiguities*)**



Exemple: 소희가 나무에 (-e) 물을 준다.(Sohee waters the trees.)

일요일에 (-e) 보자 . (See you on sunday.)

« Contextual condition 1 for the marker -e »

If the marker **-e** is followed of a binary verb of type (S\N*)\N*, **then** this marker fonctions as a determinative case suffix of dative, we assign the type of « determinative constructor » ((S\N*)/(S\N*)/N*)\N to **-e**.

- **Flexible word order**

박 장관이 파리의 모 음식점에서 김 을 만났다 .
Pak janggwani Parisuy mo eumsikjeomeso Kim-eul manassda.
Pak minister-Nom Paris-Gen restaurant-Loc Kim -Acc meet-Ps-Dc
« The minister Pak met Kim in a restaurant of Paris. »

a. N_1-i (N_2-uy N_3-eso) N_4-eul P3

N_1-Nominative (N_2-Genitive N_3-Loc) N_4-Accusative

a. N_1-i N_4-eul (N_2-uy N_3-eso) P3

b. (N_2-uy N_3-eso) N_1-i N_4-eul P3

c. (N_2-uy N_3-eso) N_4-eul N_1-i P3

d. N_4-eul N_1-i (N_2-uy N_3-eso) P3

Predicative representation

Ph a. **Ph b.** **Ph c.** **Ph d.** **Ph e.** **…..**

Categorial calculus

Introduction of combinators

a. **((B(B(C*** (-i N1) **C***(-uy N2 -eso N3)**)** **C***(-eul N4)**))** **meet)**

b. **(B(C*** (-i N1) **Bx(C***(-eul N4) **C***(-uy N2 -eso N3)**))** **meet)**

c. **(B(Bx(C***(-uy N2 -eso N3)**)** **C***(-i N1)**)** **C***(-eul N4)**)** **meet)**

d. **(Bx(B(C***(-uy N2 -eso N3)**)** **C***(-eul N4)**)** **C***(-i N1)**)** **meet)**

e. …

Reduction of the applied combinators

(((meet(-eul N_4**))** (-uy N_2 -eso N_3)) **–i** N_1**)**

- **Coordination + Positionnement croisé + Thématisation**

Ex: 장미값이 [소피는 비싸다] 고 생각하고 [폴은 저렴하다] 고 생각한다.

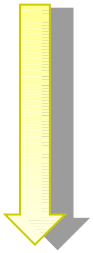
Janmigabs-**i** [Sophie-**neun** bissada]-**go** saengakha-**go** [Paul-**eun** jeoryeomhada]-**go** saenghanda

price of rose-**Th.E** Sophie-**Th.I** expensive-**Compl.** think-**Conj.** Paul-**Th.I** cheap-**Compl.** think

Sophie thinks that the price of rose is expensive and Le prix de rose, Sophie le trouve cher mais Paul le trouve bon marché.

(**C***-**i jangmigabs**) {**Φ -go {**(C*-neun Sophie)(saengakha(bissada-go))**}{**(**C***-neun Julie)(saengakhan-da(jeoryeomhada-go))**}}**

Reduction of the applied combinators

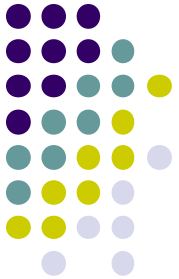


Forme Normale

-go {{((saengakha(bissada-go))-neun Sophie}-i jangmigabs}

{{((saengakhan-da(jeoryeomhada-go))-neun Julie}-i jangmigabs}

3. ACCG parser

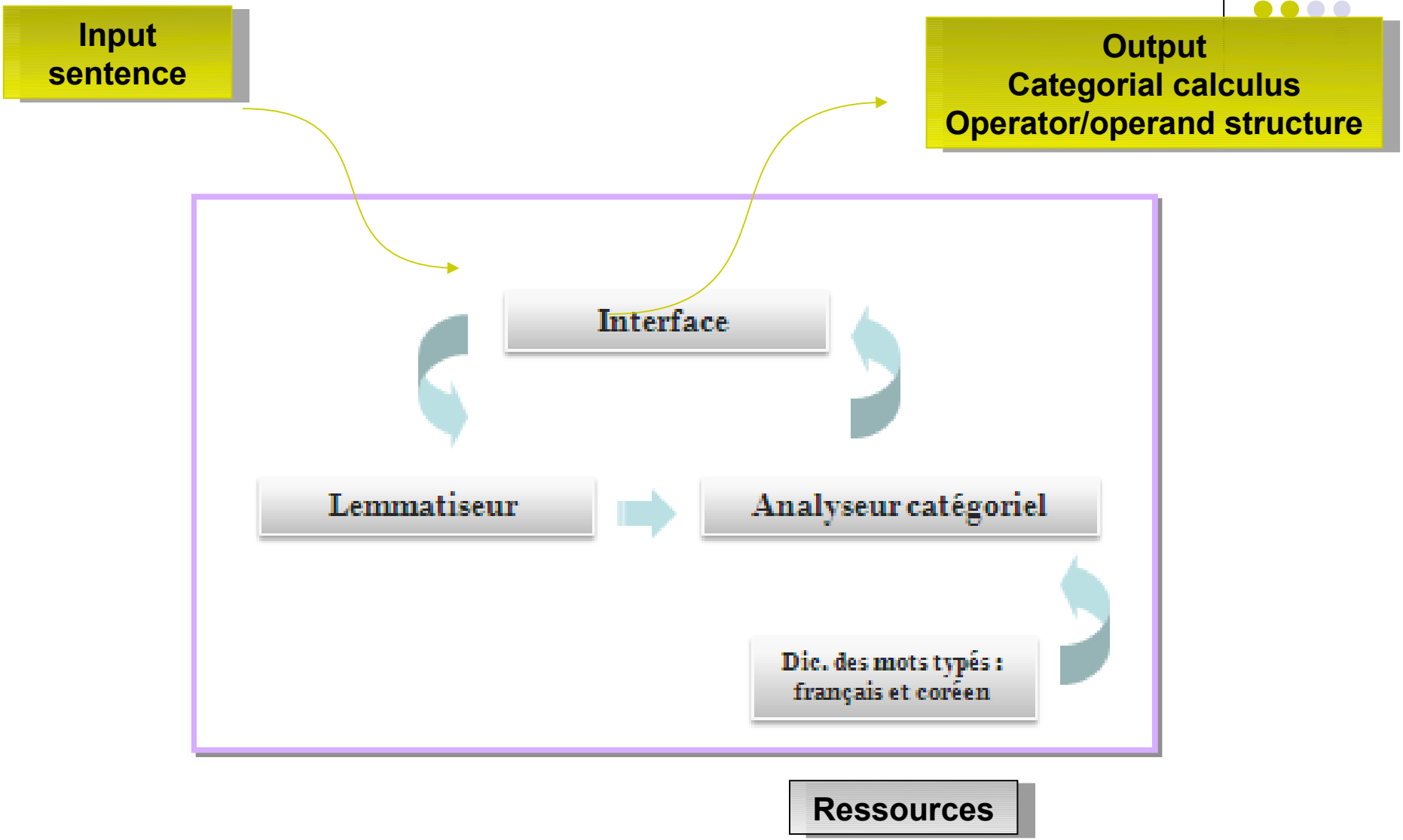


- **Objectifs**

  Develop a categorial analyzer which allows to

  - give an extended catgorial calculus
  - generate operator/operand structure
  - handle french and korean
  - take into accuount the complex linguistic phenomenon

General structure of the parser



- **Résultats comparatifs _ Comparaison de résultats(b)**
  - **Phrase2 de Texte-1: ACCG (coréen)**

지친 하루의 일을 마치고 ,

Jichin haru-uy il-eul machi-go

hard day-Gen work-Acc finish-Comp.

마리는 습관처럼 공원에서 한 시간 동안 산책을 하곤 했어요

May-neun seupgwancheoreom gongwon-eso han sigan dongan sanchaek-eul hagon hae-ss-eyo

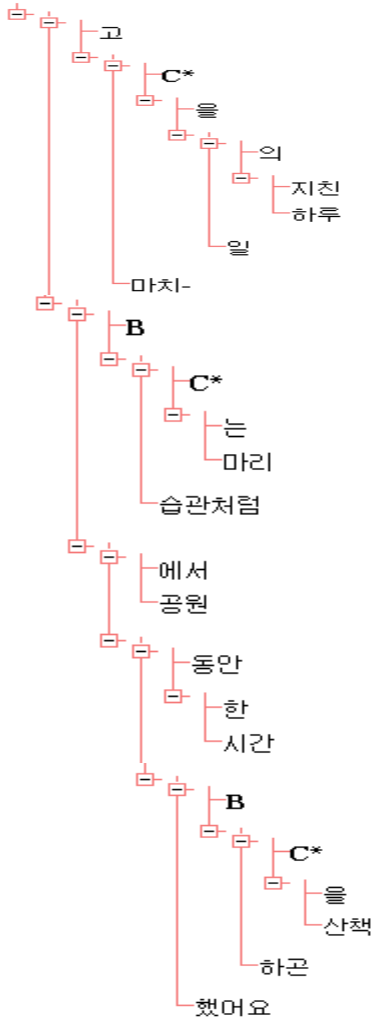
Mary-Th.I usually park-Loc an hour during working-Acc do-S.V(habit) do-Ps-Nar.

After a hard working day, Mary usually worked during an hour in the park

Calcul Catégoriel :

Type syntaxique : S

Arbre applicatif : Node (Node (Feuille= 고 , Node (Node (C* , Node (Feuille= 을 , Node (Node (Feuille= 의 , Node (Feuille= 지친 , Feuille= 하루)) , Feuille= 일))) , Feuille= 마치-)) , Node (Node (B , Node (Node (C* , Node (Feuille= 는 , Feuille= 마리)) , Feuille= 습관처럼)) , Node (Node (Feuille= 에서 , Feuille= 공원) , Node (Node (Feuille= 동안 , Node (Feuille= 한 , Feuille= 시간)) , Node (Node (B , Node (Node (C* , Node (Feuille= 을 , Feuille= 산책)) , Feuille= 하곤)) , Feuille= 했어요)))))

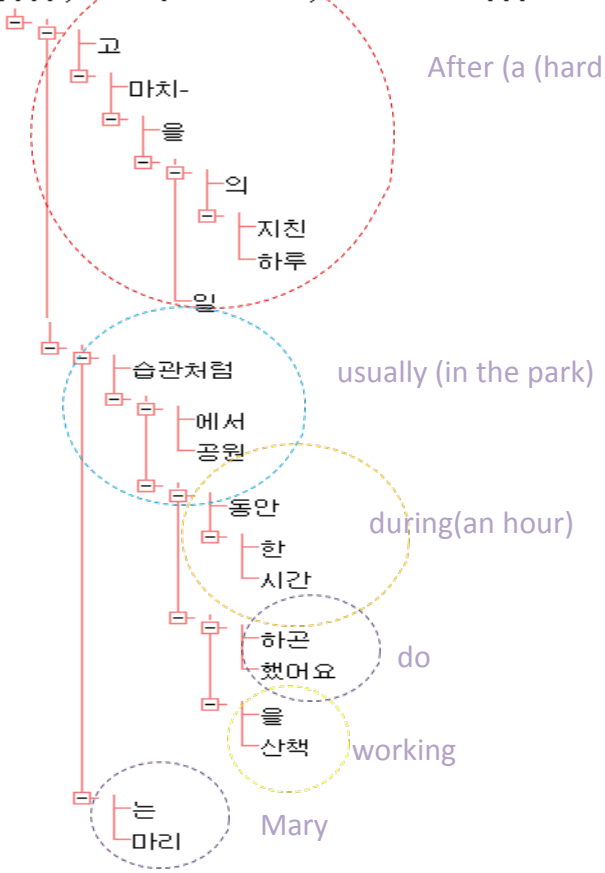


Expression combinatoire: ((고 ((C* 을 ((의 (지친 하루)) 일)) 마치-)) ((B (C* 는 마리) 습관처럼) ((에서 공원) ((동안 (한 시간)) ((B 하곤 (C* 을 산책)) 했어요))))

Structure Applicative (Forme Normale) :

Type syntaxique : S

Arbre applicatif: **Node** (**Node** (Feuille= 고 , **Node** (Feuille= 마치- , **Node** (Feuille= 을 , **Node** (**Node** (Feuille= 의 , **Node** (Feuille= 지친 , Feuille= 하루)) , Feuille= 일)))) , **Node** (**Node** (Feuille= 습관처럼 , **Node** (**Node** (Feuille= 에서 , Feuille= 공원) , **Node** (**Node** (Feuille= 동안 , **Node** (Feuille= 한 , Feuille= 시간)) , **Node** (**Node** (Feuille= 하곤 , Feuille= 했어요) , **Node** (Feuille= 을 , Feuille= 산책))))) , **Node** (Feuille= 는 , Feuille= 마리)))

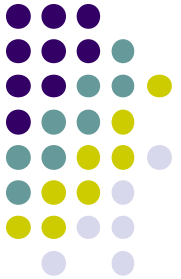


Représentation applicative: ((고 (마치- (을 ((의 (지친 하루)) 일)))) ((습관처럼 ((에서 공원) ((동안 (한 시간)) ((하곤 했어요) (을 산책))))) (는 마리))

4. Conclusion

- The ACCG can be a solution for an efficient Korean parsing by allowing to handle important morphological and syntactical problems.
- The ACCG Parser allow to generate the morpho-syntactic calculus and operator/operand structure by considering the case marker in Korean as operators.

- The parser evaluation is not still enough in order to estimate the exact parsing accuracy.
- The parser requires to use a supertagging system[Curran and Clark, 2002, 2007] for a wide-coverage.



Thank you for your attention !