

# K-SNACS: Annotating Korean Adposition Semantics

DMR 2020

Jena D. Hwang   Hanwool Choe   Na-Rae Han   Nathan Schneider



# Korean Postpositions

내 나이 여섯 살 적에, 한번은 체험담이라는 처녀림에 관한 책에서 맛있는 그림 하나를 보았다. 그것은 보아뱀 한 마리가 맹수를 삼키고 있는 그림이었다. 그걸 옮겨 놓은 그림이 위에 있다. 그 책에 이런 말이 있었다.



보아뱀은 먹이를 씹지 않고 통째로 삼킨다. 그런 다음 몸을 움직일 수가 없게 되어 먹이가 소화될 때까지 여섯 달 동안 잠을 잔다. 나는 그 그림을 보고 나서 밀림의 여러 가지 모험들을 꿈이 생각해 보았으며, 드디어는 나도 색연필을 들고 나의 첫 그림을 용케 그려 내었다.

나의 그림 제 1호, 그건 다음과 같았다.



나는 내 걸작을 어른들에게 보여주며 내 그림이 무섭지 않느냐고 물어 보았다. 어른들은 대답했다. "아니, 모자가 다 무서워?"

내 그림은 모자를 그린게 아니라 코끼리를 소화시키고 있는 보아뱀을 그린 것이었다. 그래서 나는 어른들이 알아볼 수 있도록 보아뱀의 속을 그렸다. 어른들에게만 항상 설명을 해 주어야 한다. 내 그림 제 2호는 아래와 같았다.



어른들은 나에게 속이 보였다 안 보였다 하는 보아뱀의 그림따위는 집어 치우고, 차라리 지리나 산수, 역사, 문법에 재미를 붙여 보라고 충고했다. 나는 이렇게 해서 내 나이 여섯 살 적에 화가라고 하는 맛있는 작업을 포기했다. 나는 내 그림 제 1호와 제 2호의 실패로 그만 기가 죽었던 것이다. 어른들은 자기들 혼자서는 아무것도 이해하지 못하고, 그렇다고 그 때마다 자꾸 설명을 해 주자니 어린애에게 힘든 일이었다.

그래서 나는 다른 직업을 골라야 했고, 비행기 조종을 배웠다. 나는 세계의 여기저기 제법 많은 곳을 날아다녔다. 그 덕분에 나는 한 번 속 보아도 중국과 아리조나를 구별할 수 있었다. 밤의 어둠 속에서 길을 잃을 때 지리는 매우 편리하다. 나는 이렇게 살아오는 동안 착실한 사람들을 많이 만나 자주자주 접촉을 했다. 나는 오랫동안 어른들과 함께 살며 그들을 아주 가까이서 보아왔다. 그렇다고 해서 내 의견이 크게 달라지지는 않았다.

내 나이 여섯 살 적에, 한번은 체험담이라는 처녀림에 관한 책에서 맛있는 그림 하나를 보았다. 그것은 보아뱀 한 마리가 맹수를 삼키고 있는 그림이었다. 그걸 옮겨 놓은 그림이 위에 있다. 그 책에 이런 말이 있었다.



# Korean Postpositions

- Morphological markers that mark semantic roles on nominals
- 보아뱀은 먹이를 씹지 않고 통째로 삼킨다. 그런 다음 몸을 움직일 수가 없게 되어 먹이가 소화될 때까지 여섯 달 동안 잠을 잔다. 나는 그 그림을 보고 나서 밀리의 여러 가지 모험들을 꿈이 생각해 보았으며, 드디어는 나도 색연필을 들고 나의 첫 그림을 용케 그려 내었다.

나의 그림 제 1호, 그건 다음과 같았다.



나는 내 걸작을 어른들에게 보여주며 내 그림이 무섭지 않느냐고 물어 보았다. 어른들은 대답했다. "아니, 모자가 다 무서워?"

내 그림은 모자를 그린게 아니라 코끼리를 소화시키고 있는 보아뱀을 그린 것이었다. 그래서 나는 어른들이 알아볼 수 있도록 보아뱀의 속을 그렸다. 어른들에게 항상 설명을 해 주어야 한다. 내 그림 제 2호는 아래와 같았다.



어른들은 나에게 속이 보였다 안 보였다 하는 보아뱀의 그림따위는 집어 치우고, 차라리 지리나 산수, 역사, 문법에 재미를 붙여 보라고 충고했다. 나는 이렇게 해서 내 나이 여섯 살 적에 화가라고 하는 맛있는 작업을 포기했다. 나는 내 그림 제 1호와 제 2호의 실패로 그만 기가 죽었던 것이다. 어른들은 자기를 혼자서는 아무것도 이해하지 못하고, 그렇다고 그 때마다 자꾸 설명을 해 주자니 어린애에게 힘든 일이었다.

그래서 나는 다른 직업을 골라야 했고, 비행기 조종을 배웠다. 나는 세계의 여기저기 제법 많은 곳을 날아다녔다. 그 덕분에 나는 한 번 속 보아도 중국과 아리조나를 구별할 수 있었다. 밤의 어둠 속에서 길을 잃을 때 지리는 매우 편리하다. 나는 이렇게 살아오는 동안 착실한 사람들을 많이 만나 자주자주 접촉을 했다. 나는 오랫동안 어른들과 함께 살며 그들을 아주 가까이서 보아왔다. 그렇다고 해서 내 의견이 크게 달라지지는 않았다.

내 나이 여섯 살 적에, 한번은 체험담이라는 처녀림에 관한 책에서 맛있는 그림 하나를 보았다. 그것은 보아뱀 한 마리가 맹수를 삼키고 있는 그림이었다. 그걸 옮겨 놓은 그림이 위에 있다. 그 책에 이런 말이 있었다.



# Korean Postpositions

- Morphological markers that mark semantic roles on nominals
- Highly frequent

보아뱀은 먹이를 씹지 않고 통째로 삼킨다. 그런 다음 몸을 움직일 수가 없게 되어 먹이가 소화될 때까지 여섯 달 동안 잠을 잔다. 나는 그 그림을 보고 나서 밀림의 여러 가지 모험들을 꿈이 생각해 보았으며, 드디어는 나도 색연필을 들고 나의 첫 그림을 용케 그려 내었다.

나의 그림 제 1호, 그건 다음과 같았다.



나는 내 걸작을 어른들에게 보여주며 내 그림이 무섭지 않느냐고 물어 보았다. 어른들은 대답했다. "아니, 모자가 다 무서워?"

내 그림은 모자를 그린게 아니라 코끼리를 소화시키고 있는 보아뱀을 그린 것이었다. 그래서 나는 어른들이 알아볼 수 있도록 보아뱀의 속을 그렸다. 어른들에게 항상 설명을 해 주어야 한다. 내 그림 제 2호는 아래와 같았다.



어른들은 나에게 속이 보였다 안 보였다 하는 보아뱀의 그림따위는 집어 치우고, 차라리 지리나 산수, 역사, 문법에 재미를 붙여 보라고 충고했다. 나는 이렇게 해서 내 나이 여섯 살 적에 화가라고 하는 맛있는 작업을 포기했다. 나는 내 그림 제 1호와 제 2호의 실패로 그만 기가 죽었던 것이다. 어른들은 자기들 혼자서는 아무것도 이해하지 못하고, 그렇다고 그 때마다 자꾸 설명을 해 주자니 어린애에게 힘든 일이었다.

그래서 나는 다른 직업을 골라야 했고, 비행기 조종을 배웠다. 나는 세계의 여기저기 제법 많은 곳을 날아다녔다. 그 덕분에 나는 한 번 속 보아도 중국과 아리조나를 구별할 수 있었다. 밤의 어둠 속에서 길을 잃을 때 지리는 매우 편리하다. 나는 이렇게 살아오는 동안 착실한 사람들을 많이 만나 자주자주 접촉을 했다. 나는 오랫동안 어른들과 함께 살며 그들을 아주 가까이서 보아왔다. 그렇다고 해서 내 의견이 크게 달라지지는 않았다.

내 나이에 여섯 살 적에, 한번은 체험당이라는 처녀림에 관한 책에서 맛있는 그림 하나를 보았다. 그것은 보아뱀 한 마리가 맹수 삼키고 있는 그림이었다. 그걸 옮겨 놓은 그림이 위에 있다. 그 책에 이런 말이 있었다.



보아뱀은 먹이를 씹지 않고 통째로 삼킨다. 그런 다음 몸을 움직일 수가 없게 되어 먹이가 소화될 때까지 여섯 달 동안 잠을 잔다. 나는 그 그림을 보고 나서 밀리의 여러 가지 모험들을 곰곰이 생각해 보았으며, 드디어는 나도 색연필을 들고 나의 첫 그림을 용케 그려 내었다.

나의 그림 제 1호, 그건 다음과 같았다.



나는 내 걸작을 어른들에게 보여주며 내 그림이 무섭지 않느냐고 물어 보았다. 어른들은 대답했다. "아니, 모자가 다 무서워?"

내 그림은 모자를 그린게 아니라 코끼리를 소화시키고 있는 보아뱀을 그린 것이었다. 그래서 나는 어른들이 알아볼 수 있도록 보아뱀의 속을 그렸다. 어른들<sup>에게</sup> 항상 설명을 해 주어야 한다. 내 그림 제 2호는 아래와 같았다.

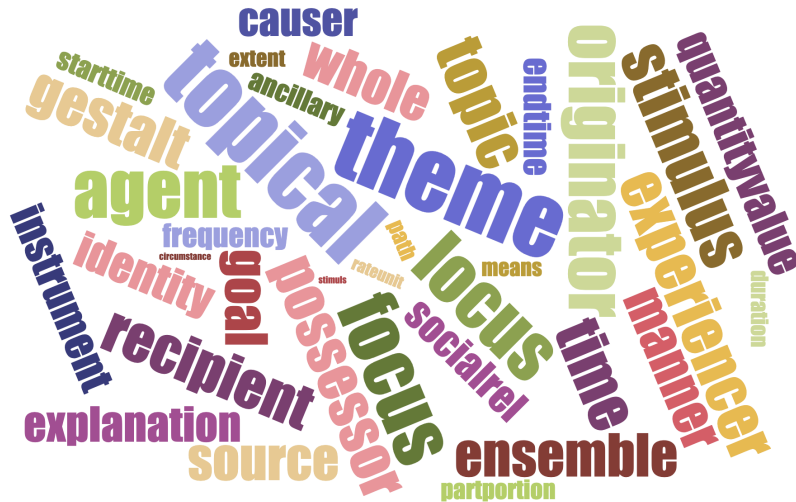


어른들은 나에게 속이 오셨다 안 오셨다 하는 보아뱀의 그림따위는 집어 치우고, 차라리 지리나 산수, 역사, 문법에 재미를 붙여 보라고 충고했다. 나는 이렇게 해서 내 나이 여섯 살 적에 화가라고 하는 맛있는 작업을 포기했다. 나는 내 그림 제 1화와 제 2화의 실패로 그만기가 죽었던 것이다. 어른들은 자기를 혼자서는 아무것도 이해하지 못하고, 그렇다고 그 때마다 자꾸 설명을 해 주자니 어린애에겐 힘겨운 일이었다.

그래서 나는 다른 직업을 골라야 했고, 비행기 조종을 배웠다. 나는 세계의 여기저기 제법 많은 곳을 날아다녔다. 그 덕분에 나는 한 번씩 보아도 중국과 아리조나를 구별할 수 있었다. 밤의 어둠 속에서 길을 잃을 때 지리는 매우 편리하다. 나는 이렇게 살아가는 동안 착실한 사람들의 많이 만나 자수자수 접촉을 했다. 나는 오랫동안 이른둥이 함께 살며 그들을 아주 가까이서 보았다. 그렇다고 해서 내 의견이 크게 달라지지는 않았다.

# Korean Postpositions

- Morphological markers that mark semantic roles on nominals
- Highly frequent
- Semantically polysemous



# Korean Postpositions

— — —

- Like English prepositions: Mediate semantics of possession, location, time, etc.

# Korean Postpositions

— — —

- Like English prepositions: Mediate semantics of possession, location, time, etc.

리사는	토름	파티에서	만났다
<b>lisa-nun</b>	<b>tom-ul</b>	<b>party-eyse</b>	mannerists
<b>Lisa-TOP</b>	<b>Tom-ACC</b>	<b>party-at</b>	met

Location

*"Lisa met Tom at the party."*

# Korean Postpositions

— — —

- Like English prepositions: Mediate semantics of possession, location, time, etc.
- Bound morphemes attaching to nominals (sub-word unit)

리사는	토름	파티에서	만났다
<b>lisa-nun</b>	<b>tom-ul</b>	<b>party-eyse</b>	mannerists
<b>Lisa-TOP</b>	<b>Tom-ACC</b>	<b>party-at</b>	met

Location

*"Lisa met Tom at the party."*



# Korean Postpositions

— — —

- Like English prepositions: Mediate semantics of possession, location, time, etc.
- Bound morphemes attaching to nominals (sub-word unit)
- Mark grammatical case (e.g., nominative (NOM), accusative (ACC))

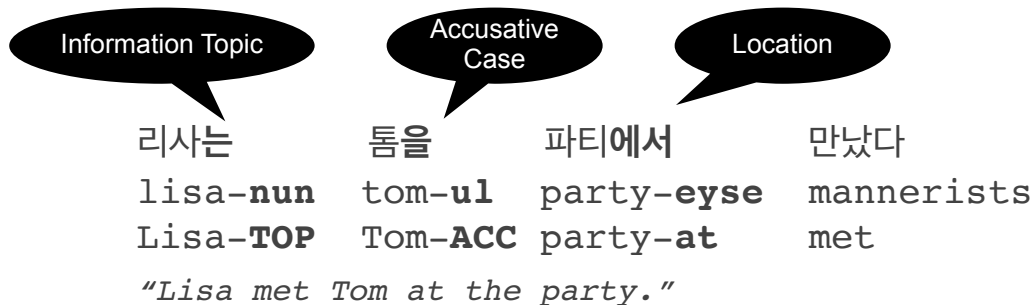
리사는	토름	파티에서	만났다
<b>lisa-nun</b>	<b>tom-ul</b>	<b>party-eyse</b>	mannerists
<b>Lisa-TOP</b>	<b>Tom-ACC</b>	<b>party-at</b>	met

*"Lisa met Tom at the party."*

# Korean Postpositions

— — —

- Like English prepositions: Mediate semantics of possession, location, time, etc.
- Bound morphemes attaching to nominals (sub-word unit)
- Mark grammatical case (e.g., nominative (NOM), accusative (ACC))
- Mediate pragmatic and discourse functions



# Introducing **K-SNACS**

Semantic **N**etwork of **A**dposition and **C**ase **S**upersenses

— — —

# Introducing **K-SNACS**

Semantic **N**etwork of **A**dposition and **C**ase **S**upersenses

— — —

- Disambiguate Korean **postposition** semantics

# Introducing **K-SNACS**

Semantic **N**etwork of **A**dposition and **C**ase **S**upersenses

- — —
- Disambiguate Korean **postposition** semantics
  - Goals
    - Corpus analysis of postpositional semantics
    - Cross-linguistic analysis
    - Automatic semantic disambiguation

# Introducing **K-SNACS**

Semantic **N**etwork of **A**dposition and **C**ase **S**upersenses

- — —
- Disambiguate Korean **postposition** semantics
  - Goals
    - Corpus analysis of postpositional semantics
    - Cross-linguistic analysis
    - Automatic semantic disambiguation
  - We adapt English SNACS

# Introducing **K-SNACS**

Semantic Network of Adposition and Case Supersenses

- — —
- Disambiguate Korean **postposition** semantics
- Goals
  - Corpus analysis of postpositional semantics
  - Cross-linguistic analysis
  - Automatic semantic disambiguation
- We adapt English SNACS
- K-SNACS is the first annotated corpus of Korean postposition supersenses

# SNACS Framework

- 50 coarse-grained semantic supersenses





# SNACS Framework

- — —
- 50 coarse-grained semantic supersenses
- Developed based on semantics of English **prepositions**
  - Annotated data: STREUSLE  
(web review corpus; Schneider et al. 2018)



# SNACS Framework

- — —
- 50 coarse-grained semantic supersenses
- Developed based on semantics of English **prepositions**
  - Annotated data: STREUSLE  
(web review corpus; Schneider et al. 2018)
- Successfully extended to Chinese Mandarin (Peng et al. 2020)



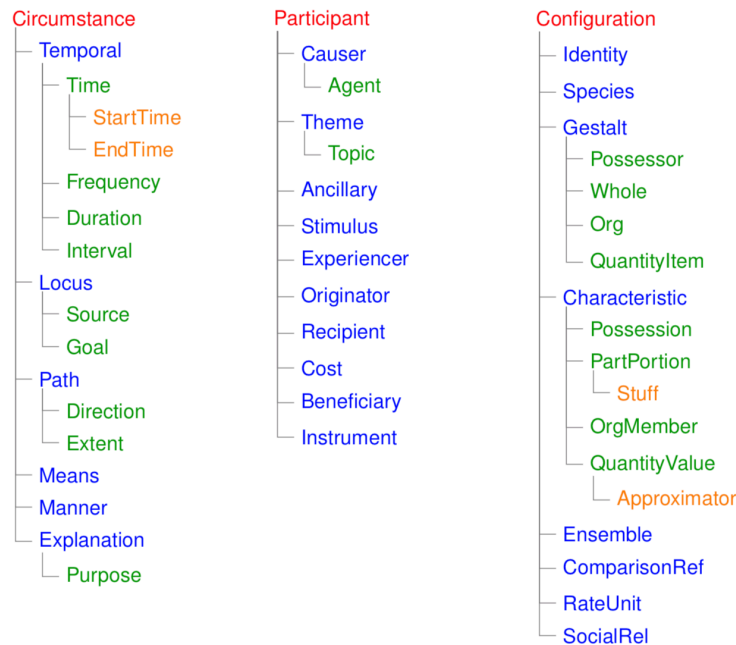
# SNACS Framework

- — —
- 50 coarse-grained semantic supersenses
- Developed based on semantics of English **prepositions**
  - Annotated data: STREUSLE  
(web review corpus; Schneider et al. 2018)
- Successfully extended to Chinese Mandarin (Peng et al. 2020)
- Preliminary works in Hindi & German



# SNACS Framework

- — —
- 50 coarse-grained semantic supersenses
- Developed based on semantics of English **prepositions**
  - Annotated data: STREUSLE  
(web review corpus; Schneider et al. 2018)
- Successfully extended to Chinese Mandarin (Peng et al. 2020)
- Preliminary works in Hindi & German



Other SNACS papers being presented at LAW:

Purpose and Accompaniment (English)

Reddit L2 English corpus

Proxy annotation tasks

# SNACS Framework

— — —

# SNACS Framework

— — —

- Semantically broad labels

# SNACS Framework

— — —

- Semantically broad labels



The cat sat **on** the mat



The cat was found **in** the box

# SNACS Framework

- Semantically broad labels



The cat sat **on** the mat

Label: **LOCUS**



The cat was found **in** the box

Label: **LOCUS**



# SNACS Framework

- Semantically broad labels that are applied not specific to any predicate



The cat **sat** on the mat

Label: **LOCUS**



The cat **was found** in the box

Label: **LOCUS**

# SNACS Framework

— — —

- Semantically broad labels that are applied not specific to any predicate
- Construal analysis:

# SNACS Framework

— — —

- Semantically broad labels that are applied not specific to any predicate
- Construal analysis:

**SCENE ROLE → FUNCTION**

# SNACS Framework

— — —

- Semantically broad labels that are applied not specific to any predicate
- Construal analysis:

**SCENE ROLE → FUNCTION**

*adposition's role with respect to the scene*

# SNACS Framework

— — —

- Semantically broad labels that are applied not specific to any predicate
- Construal analysis:

**SCENE ROLE** → **FUNCTION**

*adposition's role with respect to the scene*

*adposition's lexical meaning*

# SNACS Framework

— — —

- Semantically broad labels that are applied not specific to any predicate
- Construal analysis:

**SCENE ROLE** → **FUNCTION**

*adposition's role with respect to the scene*

*adposition's lexical meaning*



The cat sits **on** the mat

# SNACS Framework

- Semantically broad labels that are applied not specific to any predicate
- Construal analysis:

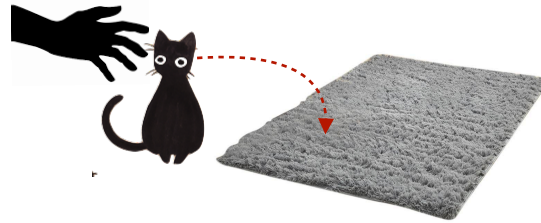
## SCENE ROLE → FUNCTION

*adposition's role with respect to the scene*

*adposition's lexical meaning*



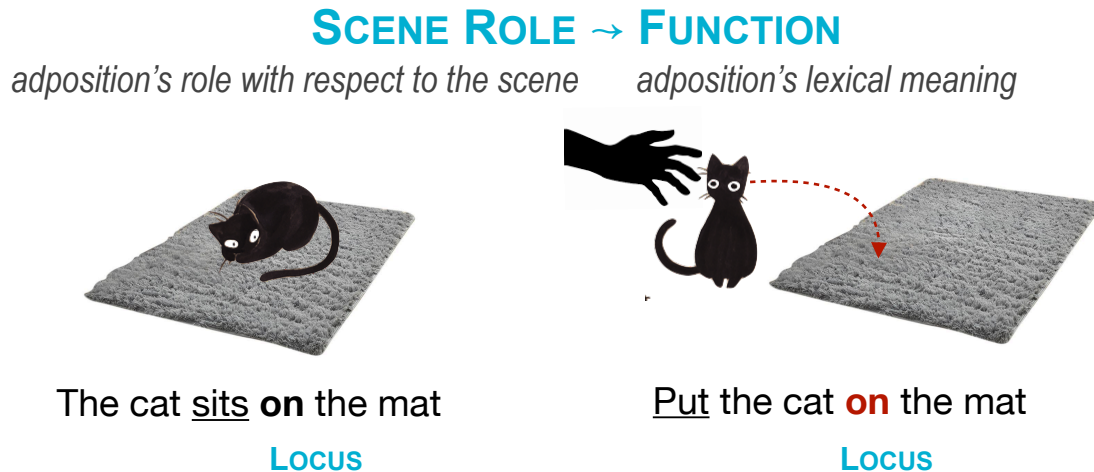
The cat sits **on** the mat



Put the cat **on** the mat

# SNACS Framework

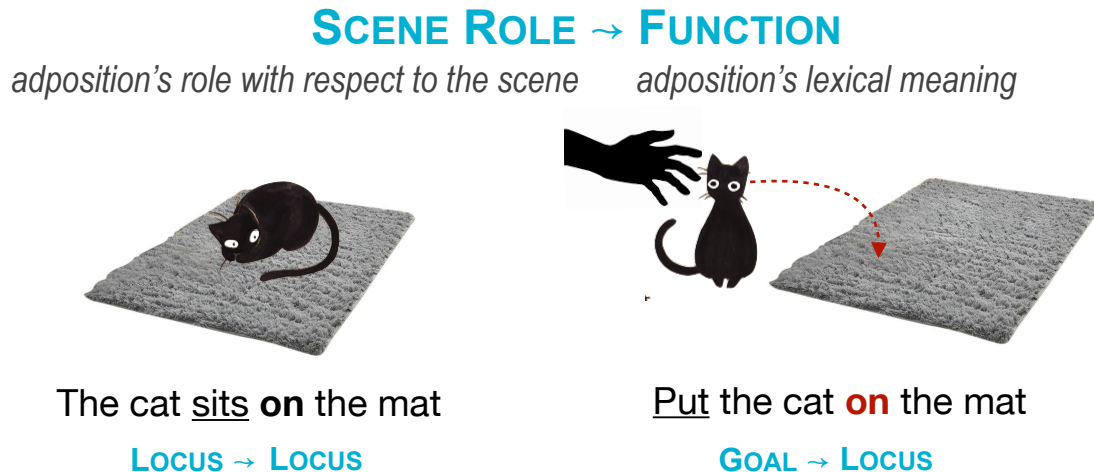
- Semantically broad labels that are applied not specific to any predicate
- Construal analysis:





# SNACS Framework

- Semantically broad labels that are applied not specific to any predicate
- Construal analysis:



# SNACS Framework

— — —

- Semantically broad labels that are applied not specific to any predicate
- Construal analysis:

**SCENE ROLE** → **FUNCTION**

*adposition's role with respect to the scene*

*adposition's lexical meaning*



The cat sits **on** the mat

**Locus** → **Locus**

# SNACS Framework

- Semantically broad labels that are applied not specific to any predicate
- Construal analysis:

## SCENE ROLE → FUNCTION

*adposition's role with respect to the scene*

*adposition's lexical meaning*



The cat sits on the mat

Locus → Locus



고양이가 매트에 앉아있다

cat mat-ey sits

# SNACS Framework

- Semantically broad labels that are applied not specific to any predicate
- Construal analysis:

## SCENE ROLE → FUNCTION

*adposition's role with respect to the scene*

*adposition's lexical meaning*



The cat sits on the mat

LOCUS → LOCUS



고양이가 매트에 앉아있다

cat mat-ey sits

LOCUS → CIRCUMSTANCE

# SNACS Framework

- Semantically broad labels that are applied not specific to any predicate
- Construal analysis:

## SCENE ROLE → FUNCTION

*adposition's role with respect to the scene*

*adposition's lexical meaning*



The cat sits on the mat

Locus → Locus



고양이가 매트에 앉아있다

cat mat-ey sits

LOCUS → CIRCUMSTANCE

# SNACS Framework

- Semantically broad labels that are applied not specific to any predicate
- Construal analysis:

## SCENE ROLE → FUNCTION

*adposition's role with respect to the scene*

*adposition's lexical meaning*



The cat sits on the mat

Locus → Locus

Locus



고양이가 매트에 앉아있다

cat mat-ey sits

LOCUS → CIRCUMSTANCE

## Applying SNACS to Korean postpositions

### General Observations

## Applying SNACS to Korean postpositions

### General Observations

- SNACS framework easily extensible to Korean postpositions



## Applying SNACS to Korean postpositions

### General Observations

- SNACS framework easily extensible to Korean postpositions
- Still there exist phenomena that SNACS can't cover

## Applying SNACS to Korean postpositions

### General Observations

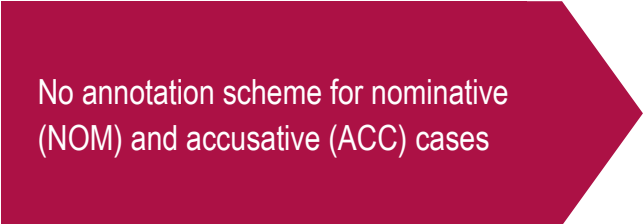
- SNACS framework easily extensible to Korean postpositions
- Still there exist phenomena that SNACS can't cover
  - We present 3 major challenges faced in the adaptation (more in the paper)

# Applying English SNACS to Korean: challenges and solutions

— — —

# Applying English SNACS to Korean: challenges and solutions

— — —



No annotation scheme for nominative (NOM) and accusative (ACC) cases

# Applying English SNACS to Korean: challenges and solutions

— — —

No annotation scheme for nominative (NOM) and accusative (ACC) cases

빌이 사과를 먹었다.

Bill-**NOM** lunch-**ACC** ate

*"Bill ate an apple"*

# Applying English SNACS to Korean: challenges and solutions

— — —

No annotation scheme for nominative (NOM) and accusative (ACC) cases

- In a transitive event, NOM and ACC markers generally align with agentivity of the participants with respect to the event (Grimm, 2011)

빌이 사과를 먹었다.  
Bill-**NOM** lunch-**ACC** ate  
*"Bill ate an apple"*

# Applying English SNACS to Korean: challenges and solutions

— — —

No annotation scheme for nominative (NOM) and accusative (ACC) cases

- In a transitive event, NOM and ACC markers generally align with agentivity of the participants with respect to the event (Grimm, 2011)
- Assign **AGENT** and **THEME** labels for NOM and ACC, respectively

빌이 **AGENT** 사과를 **THEME** 먹었다.

Bill-**NOM** lunch-**ACC** ate

*"Bill ate an apple"*

# Applying English SNACS to Korean: challenges and solutions

— — —

No annotation scheme for nominative (NOM) and accusative (ACC) cases

- In a transitive event, NOM and ACC markers generally align with agentivity of the participants with respect to the event (Grimm, 2011)
- Assign **AGENT** and **THEME** labels for NOM and ACC, respectively

빌이 **AGENT** 사과를 **THEME** 먹었다.  
Bill-**NOM** lunch-**ACC** ate  
*"Bill ate an apple"*

빌이 일찍 일어났다.  
Bill-**NOM** early woke\_up  
*"Bill woke up early"*

해가 일찍 떴다.  
sun-**NOM** early rose  
*"The sun rose early"*



# Applying English SNACS to Korean: challenges and solutions

— — —

No annotation scheme for nominative (NOM) and accusative (ACC) cases

- In a transitive event, NOM and ACC markers generally align with agentivity of the participants with respect to the event (Grimm, 2011)
- Assign **AGENT** and **THEME** labels for NOM and ACC, respectively

빌이 **AGENT** 사과를 **THEME** 먹었다.  
Bill-**NOM** lunch-**ACC** ate  
*"Bill ate an apple"*

빌이 **AGENT** 일찍 일어났다.  
Bill-**NOM** early woke\_up  
*"Bill woke up early"*

해가 **THEME** 일찍 떴다.  
sun-**NOM** early rose  
*"The sun rose early"*

# Applying English SNACS to Korean: challenges and solutions

No annotation scheme for nominative (NOM) and accusative (ACC) cases

- In a transitive event, NOM and ACC markers generally align with agentivity of the participants with respect to the event (Grimm, 2011)
- Assign **AGENT** and **THEME** labels for NOM and ACC, respectively
- If a more descriptive relationship is available given the scene, the label is assigned at the scene level, while function will retain **AGENT/THEME**

빌이 **AGENT** 사과를 **THEME** 먹었다.  
Bill-**NOM** lunch-**ACC** ate  
*"Bill ate an apple"*

빌이 **AGENT** 일찍 일어났다.  
Bill-**NOM** early woke\_up  
*"Bill woke up early"*

해가 **THEME** 일찍 떴다.  
sun-**NOM** early rose  
*"The sun rose early"*

# Applying English SNACS to Korean: challenges and solutions

No annotation scheme for nominative (NOM) and accusative (ACC) cases

- In a transitive event, NOM and ACC markers generally align with agentivity of the participants with respect to the event (Grimm, 2011)
- Assign **AGENT** and **THEME** labels for NOM and ACC, respectively
- If a more descriptive relationship is available given the scene, the label is assigned at the scene level, while function will retain **AGENT/THEME**

빌이 **AGENT** 사과를 **THEME** 먹었다.  
Bill-**NOM** lunch-**ACC** ate  
*"Bill ate an apple"*

빌이 **AGENT** 일찍 일어났다.  
Bill-**NOM** early woke\_up  
*"Bill woke up early"*

해가 **THEME** 일찍 났다.  
sun-**NOM** early rose  
*"The sun rose early"*

빌이 대답했다.  
Bill-**NOM** answered  
*"Bill answered"*

# Applying English SNACS to Korean: challenges and solutions

No annotation scheme for nominative (NOM) and accusative (ACC) cases

- In a transitive event, NOM and ACC markers generally align with agentivity of the participants with respect to the event (Grimm, 2011)
- Assign **AGENT** and **THEME** labels for NOM and ACC, respectively
- If a more descriptive relationship is available given the scene, the label is assigned at the scene level, while function will retain **AGENT/THEME**

빌이 **AGENT** 사과를 **THEME** 먹었다.  
Bill-**NOM** lunch-**ACC** ate  
*"Bill ate an apple"*

빌이 **AGENT** 일찍 일어났다.  
Bill-**NOM** early woke\_up  
*"Bill woke up early"*

해가 **THEME** 일찍 났다.  
sun-**NOM** early rose  
*"The sun rose early"*

빌이 **ORIGINATOR→AGENT** 대답했다.  
Bill-**NOM** answered  
*"Bill answered"*

# Applying English SNACS to Korean: challenges and solutions

— — —

# Applying English SNACS to Korean: challenges and solutions

— — —



SNACS does not cover pragmatic uses

# Applying English SNACS to Korean: challenges and solutions

— — —

SNACS does not cover pragmatic uses

- Korean postpositions mark a range of pragmatic usages

# Applying English SNACS to Korean: challenges and solutions

— — —

SNACS does not cover pragmatic uses

- Korean postpositions mark a range of pragmatic usages

빌은            일을            잘            했다  
Bill-**TOP**   work-ACC   well   did  
*"As for Bill,  
he did a good job"*

빌**까지**            일을            잘            했다  
Bill-**even**        work-ACC   well   did  
*"Even Bill (the least likely  
candidate) did a good job"*



# Applying English SNACS to Korean: challenges and solutions

— — —

SNACS does not cover pragmatic uses

- Korean postpositions mark a range of pragmatic usages

빌은            일을        잘        했다  
Bill-**TOP**   work-ACC   well   did  
*"As for Bill,  
he did a good job"*

빌**까지**            일을        잘        했다  
Bill-**even**        work-ACC   well   did  
*"Even Bill (the least likely  
candidate) did a good job"*

Semantic meaning: "Bill did a good job."

# Applying English SNACS to Korean: challenges and solutions

— — —

SNACS does not cover pragmatic uses

- Korean postpositions mark a range of pragmatic usages
- We create two new labels **TOPICAL** and **FOCUS**

빌은            일을            잘            했다  
Bill-**TOP**    work-ACC    well    did  
*"As for Bill,  
he did a good job"*

빌**까지**            일을            잘            했다  
Bill-**even**        work-ACC    well    did  
*"Even Bill (the least likely  
candidate) did a good job"*

Semantic meaning: "Bill did a good job."

# Applying English SNACS to Korean: challenges and solutions

— — —

SNACS does not cover pragmatic uses

- Korean postpositions mark a range of pragmatic usages
- We create two new labels **TOPICAL** and **FOCUS**
  - Information topic marker receives **TOPICAL** label

빌은 **TOPICAL** 일을 잘 했다  
Bill-**TOP** work-ACC well did  
*"As for Bill,  
he did a good job"*

빌**까지** 일을 잘 했다  
Bill-**even** work-ACC well did  
*"Even Bill (the least likely  
candidate) did a good job"*

Semantic meaning: "Bill did a good job."

# Applying English SNACS to Korean: challenges and solutions

— — —

SNACS does not cover pragmatic uses

- Korean postpositions mark a range of pragmatic usages
- We create two new labels **TOPICAL** and **FOCUS**
  - Information topic marker receives **TOPICAL** label
  - Focus markers receive **FOCUS** label

빌은 **TOPICAL** 일을 잘 했다  
Bill-**TOP** work-ACC well did  
*"As for Bill,  
he did a good job"*

빌까지 **FOCUS** 일을 잘 했다  
Bill-**even** work-ACC well did  
*"Even Bill (the least likely  
candidate) did a good job"*

Semantic meaning: "Bill did a good job."

# Applying English SNACS to Korean: challenges and solutions

SNACS does not cover pragmatics

## Circumstance

- Temporal
  - Time
    - StartTime
    - EndTime
  - Frequency
  - Duration
  - Interval
- Locus
  - Source
  - Goal
- Path
  - Direction
  - Extent
- Means
- Manner
- Explanation
  - Purpose

## Participant

- Causer
  - Agent
- Theme
  - Topic
  - Quote
- Stimulus
- Experiencer
- Originator
- Recipient
- Cost
- Beneficiary
- Instrument

## Configuration

- Identity
- Species
- Gestalt
  - Possessor
  - Whole
  - Org
  - QuantityItem
- Characteristic
  - Possession
  - PartPortion
    - Stuff
  - QuantityValue
    - Approximator
  - OrgMember
- ComparisonRef
- RateUnit
- SocialRel
- Ancillary
- Ensemble

## Context

- Focus
- Topical

2 new labels under  
**CONTEXT** the new  
pragmatics branch


빌은 **TOPICAL** 일을 잘  
Bill-**TOP** work-ACC we.  
"As for Bill,  
he did a good job"

# Applying English SNACS to Korean: challenges and solutions

— — —

# Applying English SNACS to Korean: challenges and solutions

— — —



No good method for handling  
quotative postpositions

# Applying English SNACS to Korean: challenges and solutions

— — —

No good method for handling  
quotative postpositions

- Korean Quotative postpositions mark the hearsay *participant* arguments of verbs of communication and cogitation



# Applying English SNACS to Korean: challenges and solutions

— — —

No good method for handling  
quotative postpositions

- Korean Quotative postpositions mark the hearsay *participant* arguments of verbs of communication and cogitation

빌은            오늘    도착한다<sup>고</sup>            했다  
Bill-TOP today arrive-**quote**            said  
"[They] said Bill would arrive today"

# Applying English SNACS to Korean: challenges and solutions

— — —

No good method for handling  
quotative postpositions

- Korean Quotative postpositions mark the hearsay *participant* arguments of verbs of communication and cogitation
- We introduce a new **QUOTE** label

빌은            오늘    도착한다<sup>고</sup>            했다  
Bill-TOP today arrive-**quote**            said  
“[They] said Bill would arrive today”

# Applying English SNACS to Korean: challenges and solutions

— — —

No good method for handling  
quotative postpositions

- Korean Quotative postpositions mark the hearsay *participant* arguments of verbs of communication and cogitation
- We introduce a new **QUOTE** label
  - currently used only as at the function level

빌은            오늘    도착한다**고**            했다  
Bill-TOP today arrive-**quote**            said  
“[They] said Bill would arrive today”

# Applying English SNACS to Korean: challenges and solutions

— — —

No good method for handling  
quotative postpositions

- Korean Quotative postpositions mark the hearsay *participant* arguments of verbs of communication and cogitation
- We introduce a new **QUOTE** label
  - currently used only as at the function level

빌은            오늘    도착한다 **고** **TOPIC→QUOTE** 했다  
Bill-TOP today arrive-**quote**            said  
"[They] said Bill would arrive today"

# Applying English SNACS to Korean: challenges and solutions

No good method for handling  
quotative postpositions



## K-SNACS Corpus of The Little Prince

# The Korean *Little Prince* Corpus

Raw data

consisting of  
**27 chapters**  
with **10,939 tokens**



# The Korean *Little Prince* Corpus

Raw data

Target postpositions

consisting of  
**27 chapters**  
with **10,939 tokens**

automatic tokenization  
with KOMA tagger  
(Lee and Rim 2009)

**4,166 postpositional  
tokens**





# The Korean *Little Prince* Corpus

## Raw data

consisting of  
**27 chapters**  
with **10,939 tokens**



## Target postpositions

automatic tokenization  
with KOMA tagger  
(Lee and Rim 2009)

**4,166 postpositional  
tokens**

## Guidelines

Three chapters used to  
establish K-SNACS  
guidelines

**Korean guidelines at:**  
[github.com/jdch00/k-snacs](https://github.com/jdch00/k-snacs)

# The Korean *Little Prince* Corpus

## Raw data

consisting of  
**27 chapters**  
with **10,939 tokens**



## Target postpositions

automatic tokenization  
with KOMA tagger  
(Lee and Rim 2009)

**4,166 postpositional  
tokens**

## Guidelines

Three chapters used to  
establish K-SNACS  
guidelines

**Korean guidelines at:  
[github.com/jdch00/k-snacs](https://github.com/jdch00/k-snacs)**

## Annotation

The remaining chapters  
are double annotated and  
adjudicated by consensus

**IAA: 82.7% @ scene role  
90.2% @ functional label**

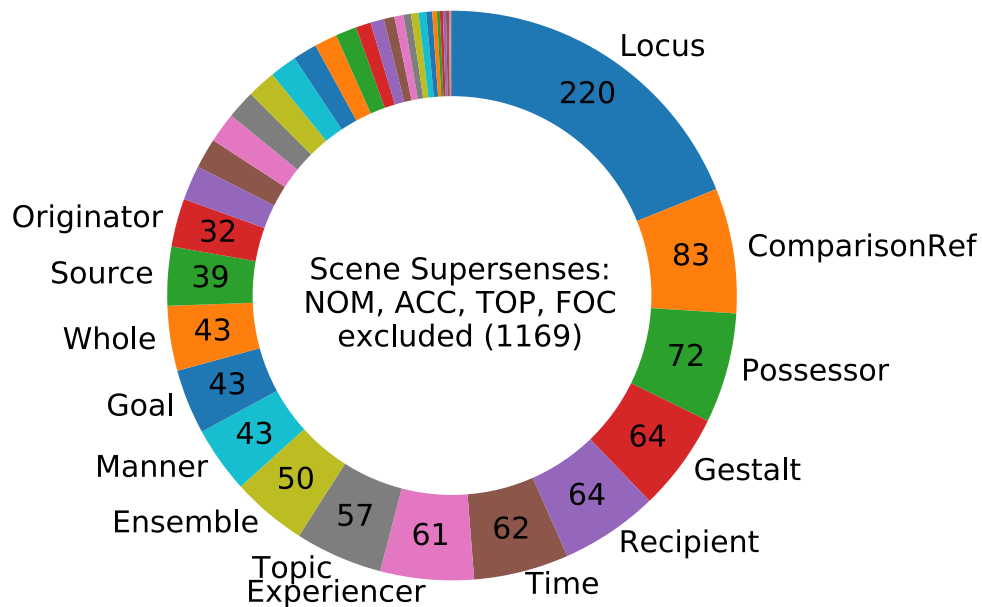
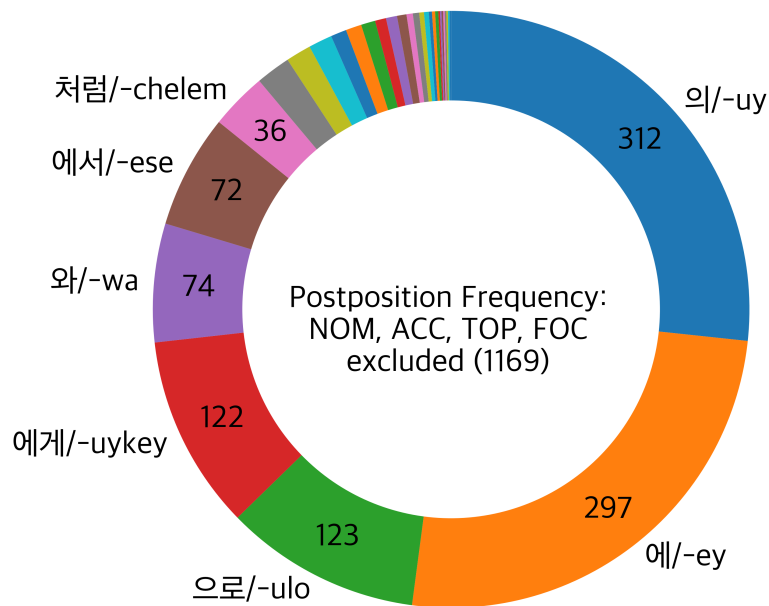
# Distribution: Postpositions vs. Scene Labels

— — —

- NOM, ACC, FOC, TOP postpositions account for ~70% of all annotation targets

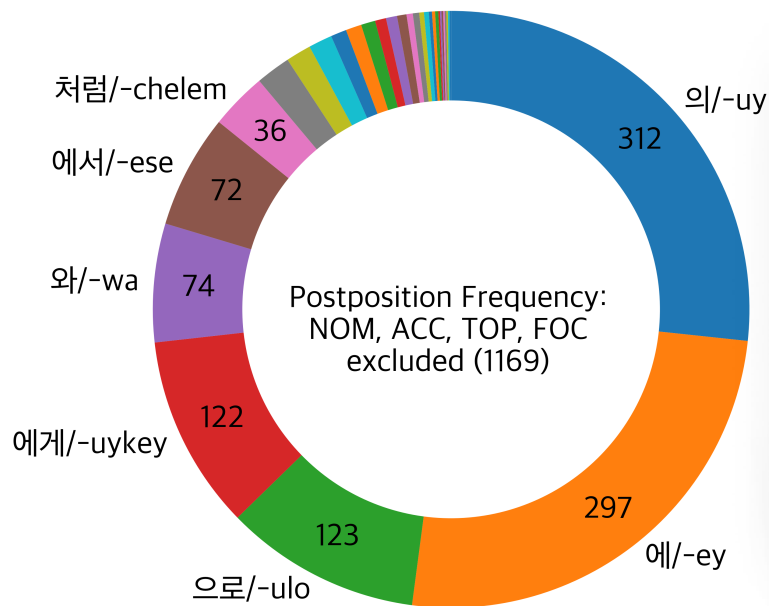
# Distribution: Postpositions vs. Scene Labels

- NOM, ACC, FOC, TOP postpositions account for ~70% of all annotation targets



# Distribution: Postpositions vs. Scene Labels

- NOM, ACC, FOC, TOP postpositions account for ~70% of all annotation targets

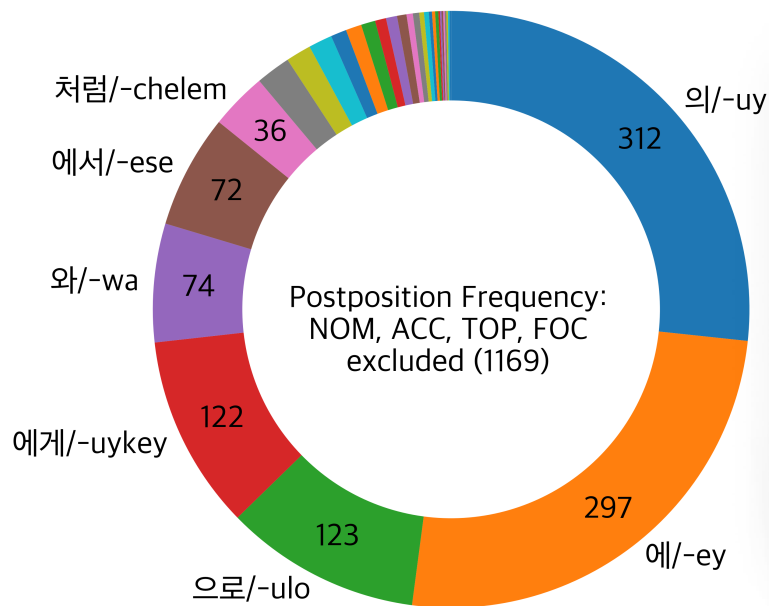


- 의/-uy prototypically genitive case (**POSSESSOR**)  
↳ total of 18 distinct scene roles! (**CHARACTERISTIC** and **QUANTITYVALUE**)

Topic  
Experiencer Time

# Distribution: Postpositions vs. Scene Labels

- NOM, ACC, FOC, TOP postpositions account for ~70% of all annotation targets

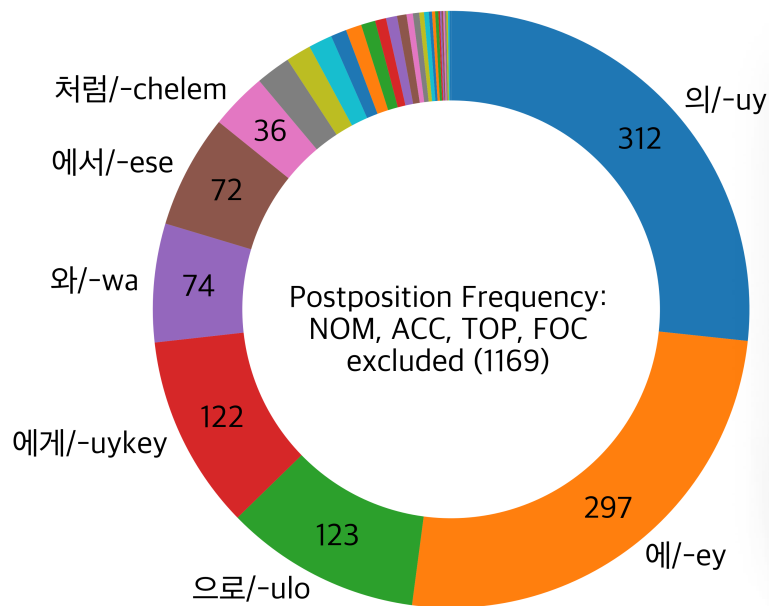


- 의/-uy prototypically genitive case (**POSSESSOR**)  
↳ total of 18 distinct scene roles! (**CHARACTERISTIC** and **QUANTITYVALUE**)
- 에/-ey bleached postposition (**CIRCUMSTANCE**)  
↳ total of 16 unique scene roles

Topic  
Experiencer Time

# Distribution: Postpositions vs. Scene Labels

- NOM, ACC, FOC, TOP postpositions account for ~70% of all annotation targets

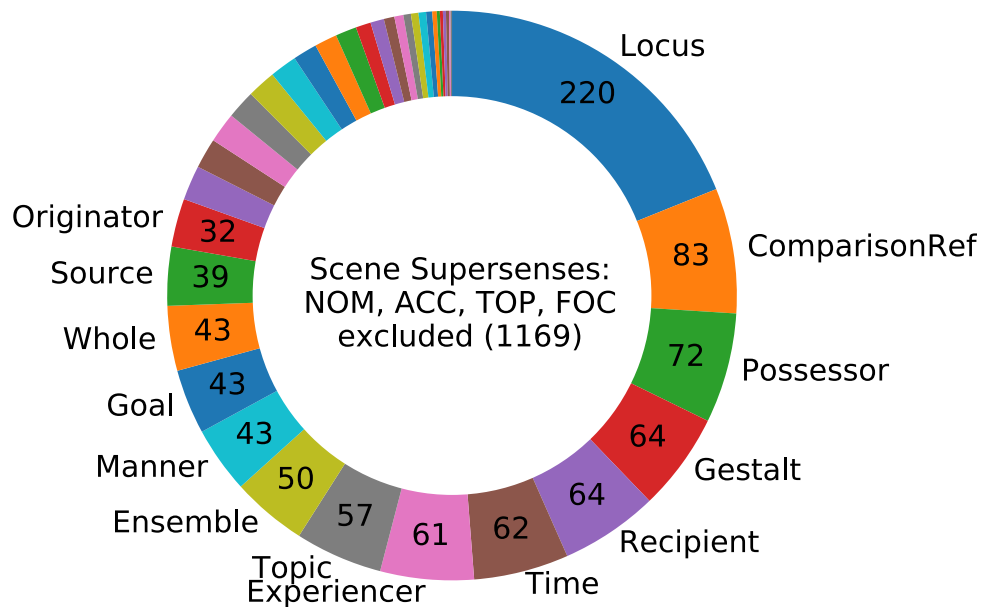
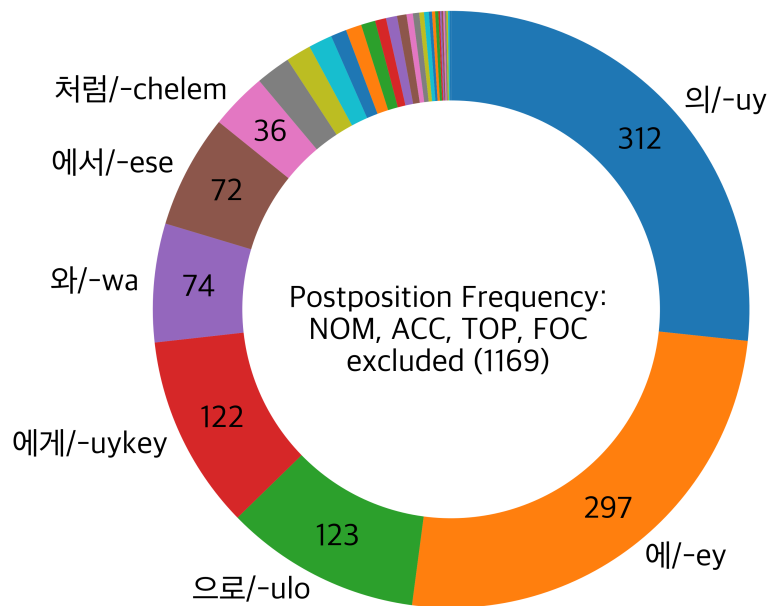


- **의/-uy** prototypically genitive case (**POSSESSOR**)  
↳ total of 18 distinct scene roles! (**CHARACTERISTIC** and **QUANTITYVALUE**)
- **에/-ey** bleached postposition (**CIRCUMSTANCE**)  
↳ total of 16 unique scene roles
- **으로/-ulo** prototypically lative case (**DIRECTION**)  
↳ total of 15 unique scene roles (**INSTRUMENT**, **IDENTITY**, and **GOAL**)

Topic  
Experiencer Time

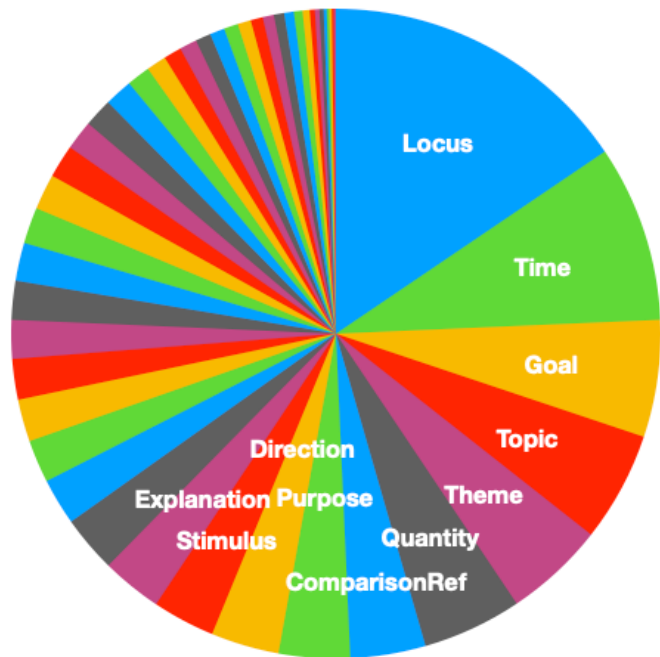
# Distribution: Postpositions vs. Scene Labels

- NOM, ACC, FOC, TOP postpositions account for ~70% of all annotation targets



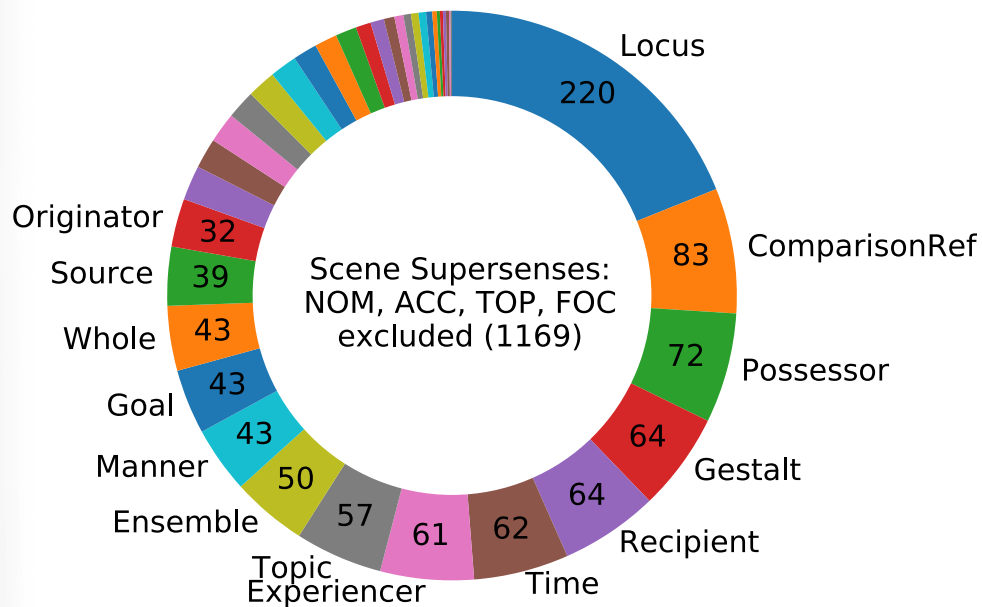


# Distribution: Postpositions vs. Scene Labels



Scene Label distribution in English STREUSLE  
(Schneider et al. 2018)

t for ~70% of all annotation targets

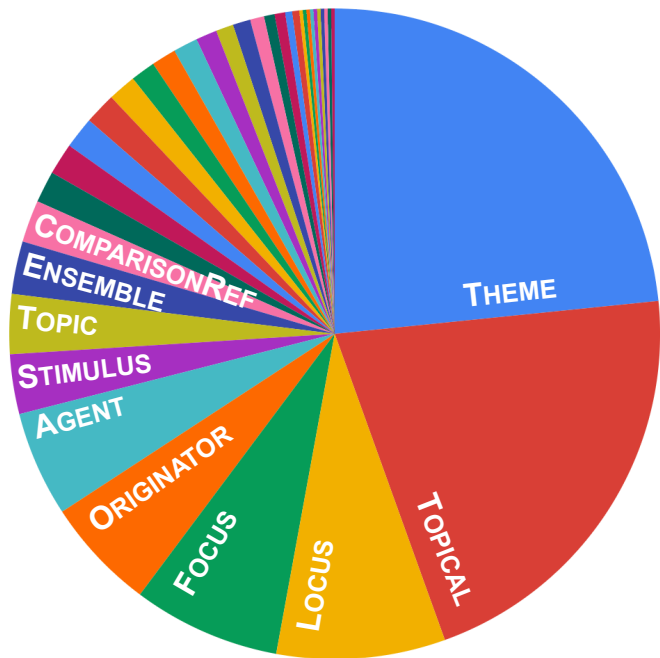


# Ko vs. En: Little Prince Chapters 1-7

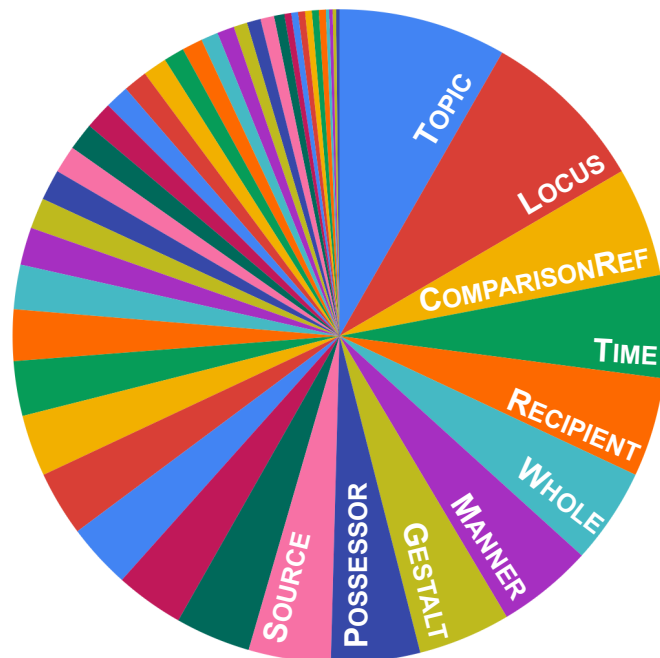
— — —

# Ko vs. En: Little Prince Chapters 1-7

--

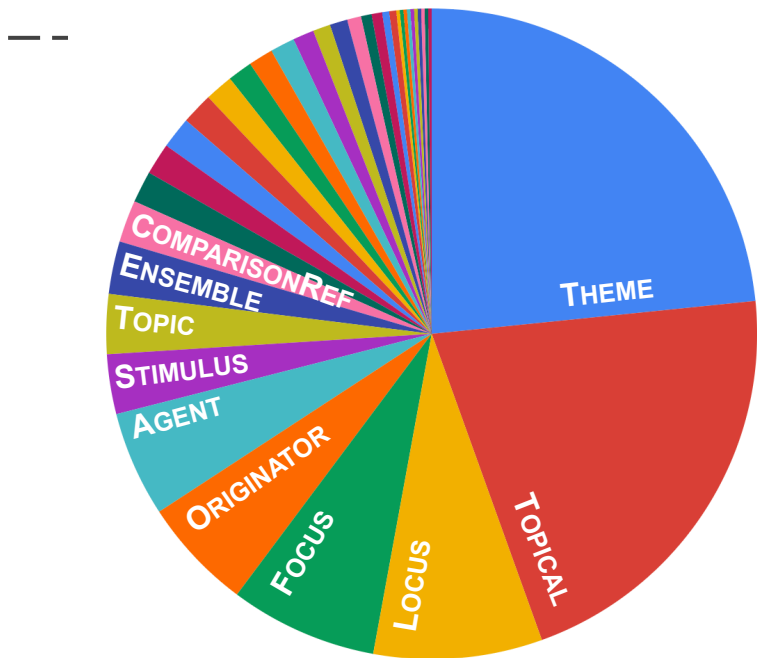


The Little Prince - KOREAN  
1144 tokens, 36 scene roles

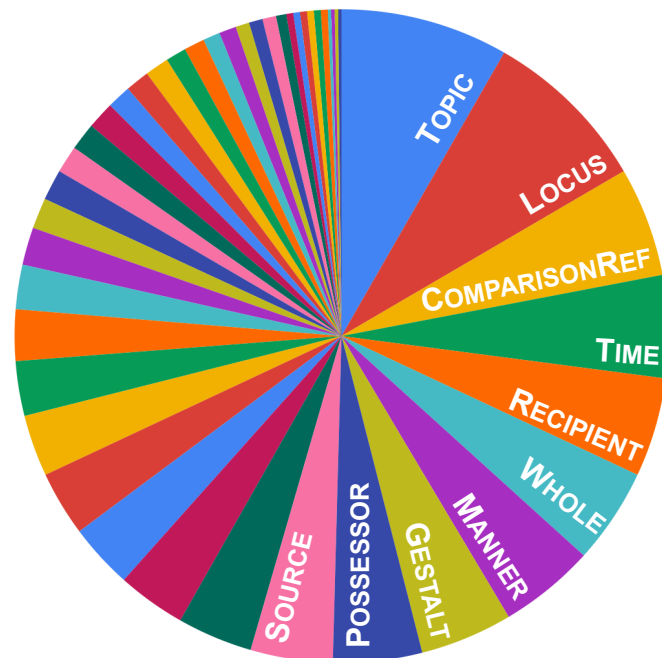


The Little Prince - ENGLISH  
591 tokens, 45 scene roles

# Ko vs. En: Little Prince Chapters 1-7



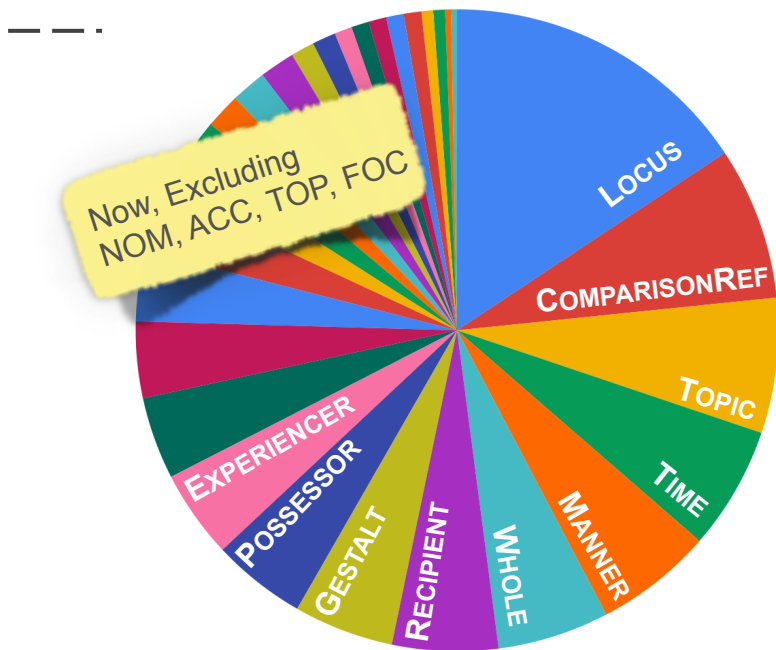
The Little Prince - KOREAN  
1144 tokens, 36 scene roles



The Little Prince - ENGLISH  
591 tokens, 45 scene roles

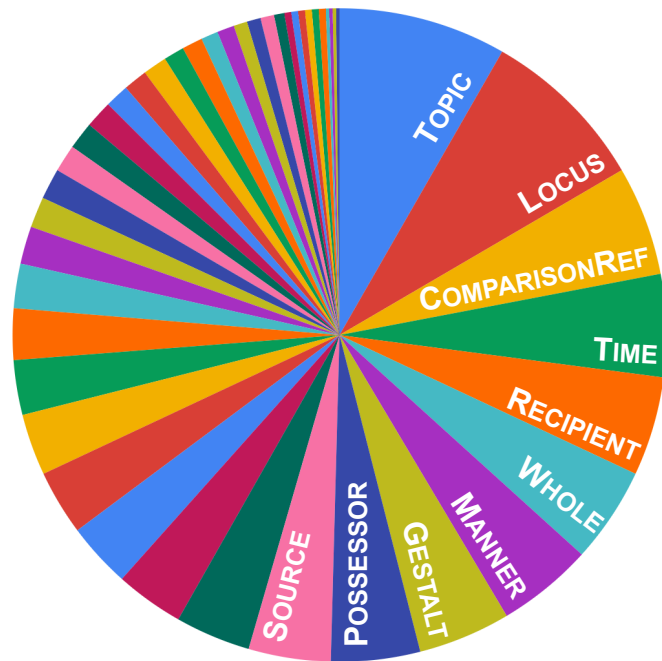
*more tokens &  
fewer SNACS types*

# Ko vs. En: Little Prince Chapters 1-7



The Little Prince - KOREAN  
340 tokens, 31 scene roles

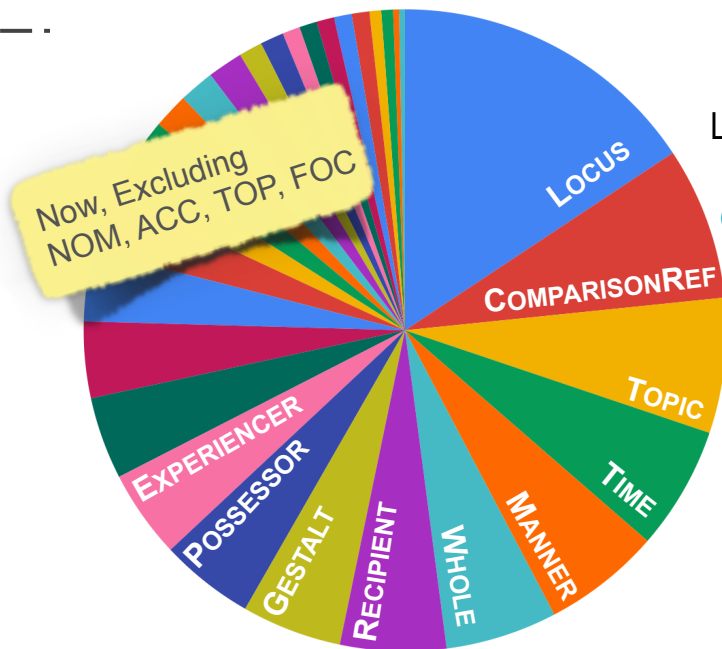
*note: the discussion in paper only removes  
NOM, ACC, TOP postpositions*



The Little Prince - ENGLISH  
591 tokens, 45 scene roles

# Ko vs. En: Little Prince Chapters 1-7

— — .

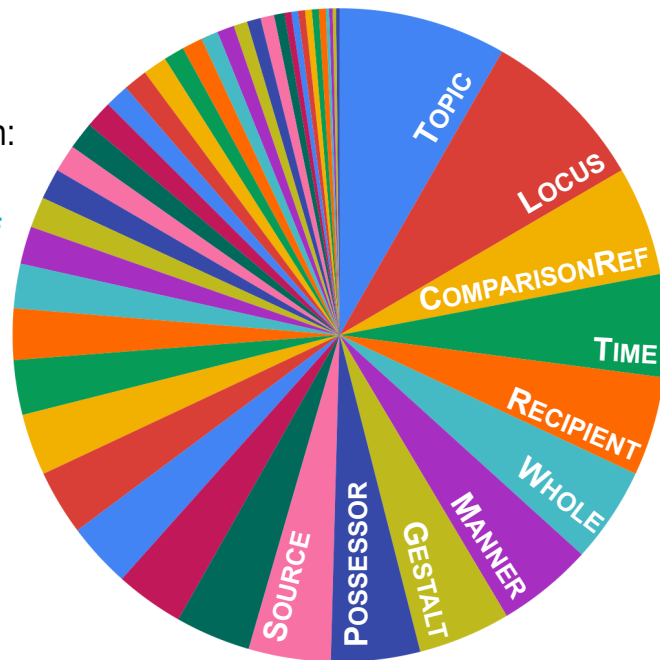


The Little Prince - KOREAN  
340 tokens, 31 scene roles

*note: the discussion in paper only removes  
NOM, ACC, TOP postpositions*

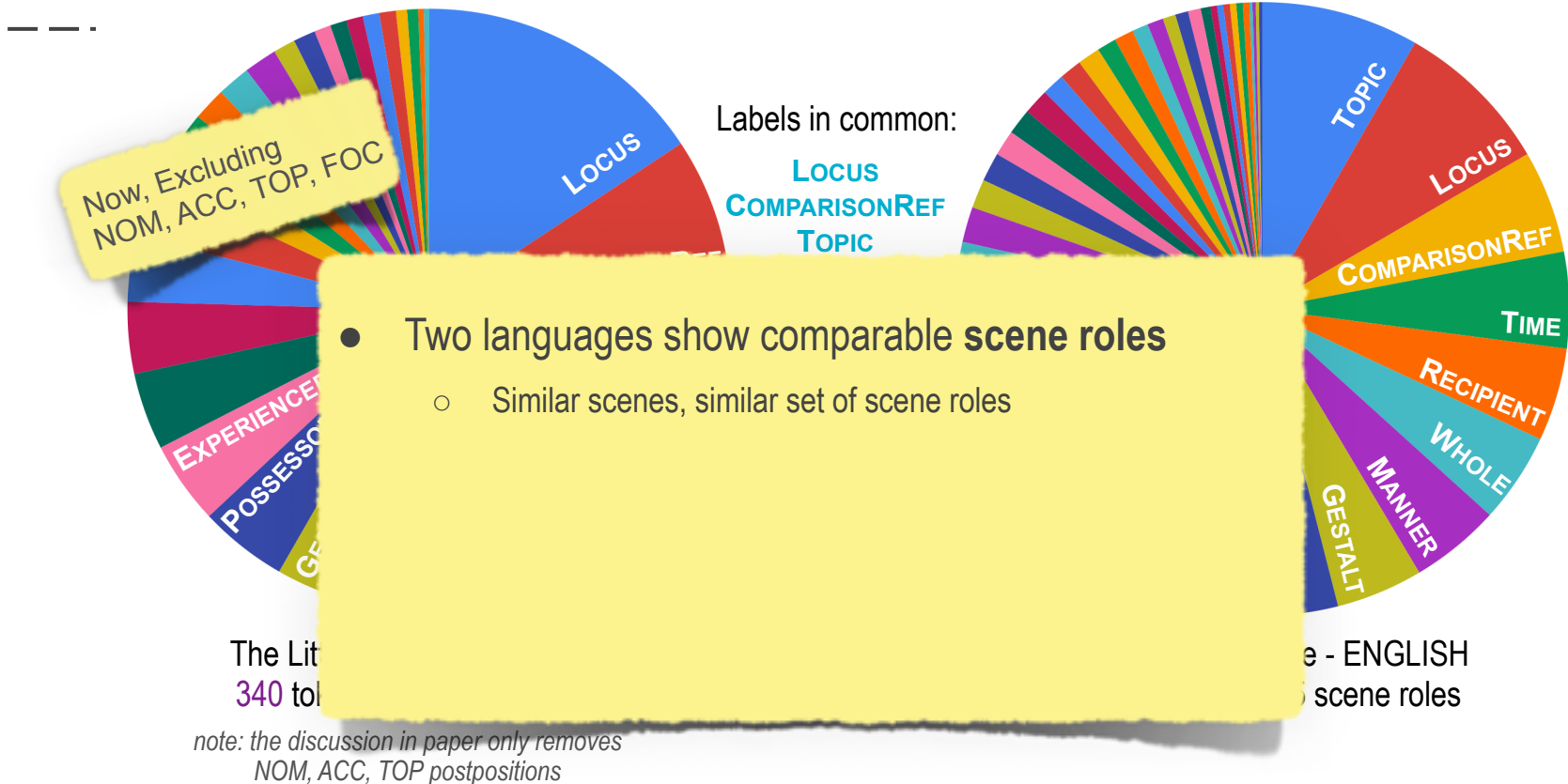
Labels in common:

LOCUS  
COMPARISONREF  
TOPIC  
TIME  
MANNER  
WHOLE  
RECIPIENT  
GESTALT  
POSSESSOR

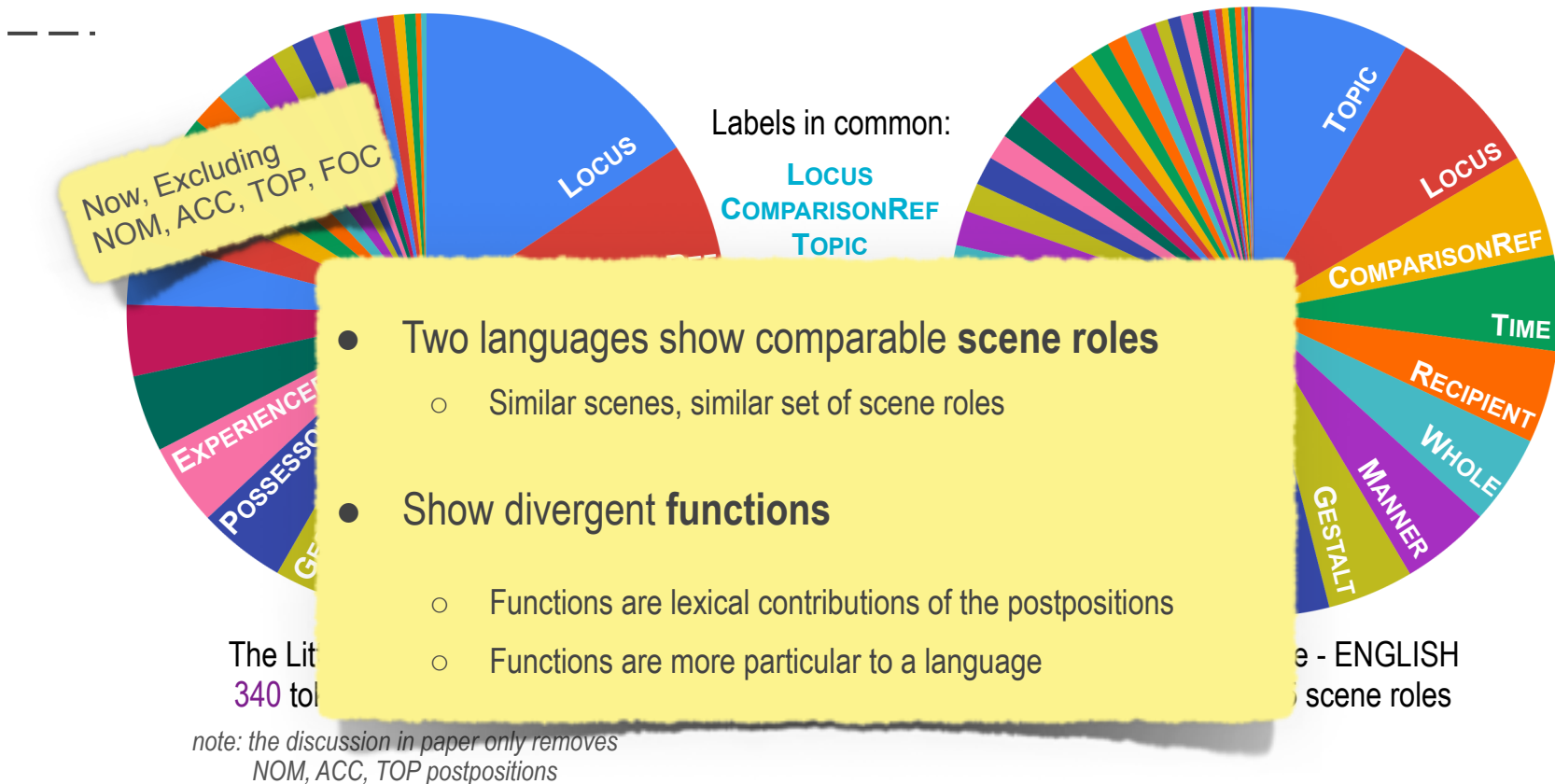


The Little Prince - ENGLISH  
591 tokens, 45 scene roles

# Ko vs. En: Little Prince Chapters 1-7



# Ko vs. En: Little Prince Chapters 1-7





# Alignment study (Chapters 2 and 7)

— — —

# Alignment study (Chapters 2 and 7)

— — —

- Adposition to adposition alignment only

# Alignment study (Chapters 2 and 7)

— — —

- Adposition to adposition alignment only
- Expectation:
  - High scene agreement (same story, similar scenes)
  - Lower function agreement (language specific choice)

# Alignment study (Chapters 2 and 7)

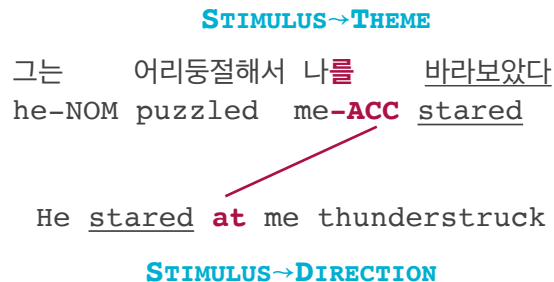
— — —

- Adposition to adposition alignment only
- Expectation:
  - High scene agreement (same story, similar scenes)
  - Lower function agreement (language specific choice)
  - Disagreements: same scene, different function

# Alignment study (Chapters 2 and 7)

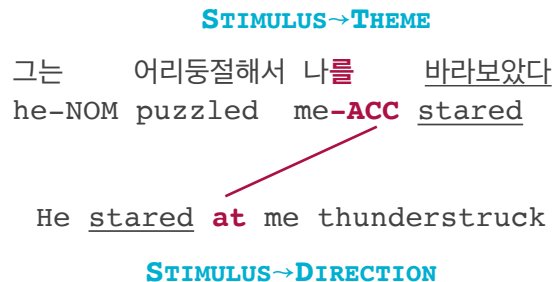
— — —

- Adposition to adposition alignment only
- Expectation:
  - High scene agreement (same story, similar scenes)
  - Lower function agreement (language specific choice)
  - Disagreements: same scene, different function



# Alignment study (Chapters 2 and 7)

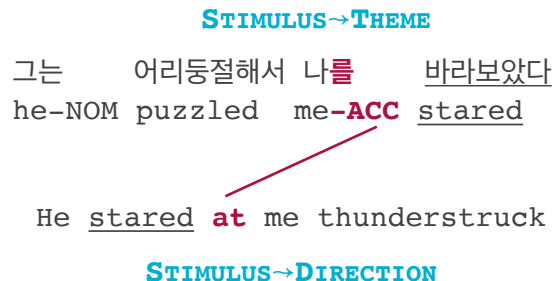
- Adposition to adposition alignment only
- Expectation:
  - High scene agreement (same story, similar scenes)
  - Lower function agreement (language specific choice)
  - Disagreements: same scene, different function
- Observations:



# Alignment study (Chapters 2 and 7)

— — —

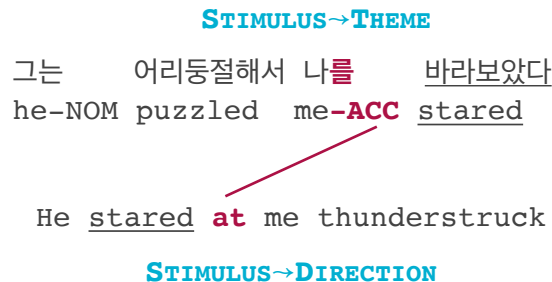
- Adposition to adposition alignment only
- Expectation:
  - High scene agreement (same story, similar scenes)
  - Lower function agreement (language specific choice)
  - Disagreements: same scene, different function
- Observations:
  - Among 226 English and 410 Korean adpositions tokens  
**81 aligned tokens**



# Alignment study (Chapters 2 and 7)

— — —

- Adposition to adposition alignment only
- Expectation:
  - High scene agreement (same story, similar scenes)
  - Lower function agreement (language specific choice)
  - Disagreements: same scene, different function
- Observations:
  - Among 226 English and 410 Korean adpositions tokens  
**81 aligned tokens**
  - Agreement lower than expected  
**scene role: 66.7%**  
**function label: 38.3%**





# Alignment study (Chapters 2 and 7)

- Adposition to adposition alignment only
- Expectation:
  - High scene agreement (same story, similar scenes)
  - Lower function agreement (language specific choice)
  - Disagreements: same scene, different function
- Observations:
  - Among 226 English and 410 Korean adpositions tokens  
**81 aligned tokens**
  - Agreement lower than expected  
**scene role: 66.7%**  
**function label: 38.3%**

**STIMULUS→THEME**

그는 어리둥절해서 나를 바라보았다  
he-NOM puzzled me-ACC stared

He stared at me thunderstruck

**STIMULUS→DIRECTION**



Why?!

# Alignment study (Chapters 2 and 7)

— — —

- Q1: *Why is the alignment so low?*

# Alignment study (Chapters 2 and 7)

— — —

- Q1: *Why is the alignment so low?*

EN: I jumped **to** my feet

# Alignment study (Chapters 2 and 7)

— — —

- Q1: *Why is the alignment so low?*

벌떡            일어섰다    “*[I] stood up suddenly*”  
suddenly   stood-up

EN: I jumped **to** my feet

# Alignment study (Chapters 2 and 7)

— — —

- Q1: *Why is the alignment so low?*
  - Idiomatic expressions

벌떡            일어섰다    “*[I] stood up suddenly*”  
suddenly   stood-up

EN: I jumped **to** my feet

# Alignment study (Chapters 2 and 7)

- Q1: Why is the alignment so low?

- Idiomatic expressions

벌떡      일어섰다  
suddenly stood-up

“[I] stood up suddenly”

EN: I jumped **to** my feet

나는    이렇게 해서    진심을    털어 놓고 이야기 할 사람도    없이    혼자    살아왔다  
I-NOM   so   (it)was heart-ACC   confide speak do person-FOC   lacking   alone   lived

“So it was that I lived alone lacking people to speak to and confide my genuine feelings”

EN: So I lived **my** life alone **without** anyone that I could really talk **to**

# Alignment study (Chapters 2 and 7)

- Q1: Why is the alignment so low?

- Idiomatic expressions
- Cross linguistic differences in expression

벌떡            일어섰다    “[[I] stood up suddenly”  
suddenly stood-up

EN: I jumped **to** my feet

나는    이렇게 해서    진심을    털어 놓고 이야기 할 사람도    없이    혼자    살아왔다  
I-NOM   so   (it)was heart-ACC   confide speak do person-FOC   lacking   alone   lived

“So it was that I lived  
alone lacking people to  
speak to and confide my  
genuine feelings”

EN: So I lived **my** life alone **without** anyone that I could really talk **to**

# Alignment study (Chapters 2 and 7)

- Q1: Why is the alignment so low?

- Idiomatic expressions
- Cross linguistic differences in expression

벌떡 일어섰다 "[] stood up suddenly"  
suddenly stood-up

EN: I jumped **to** my feet

나는 이렇게 해서 진심을 털어 놓고 이야기 할 사람도 없이 혼자 살아왔다  
I-NOM so (it)was heart-ACC confide speak do person-FOC lacking alone lived

"So it was that I lived alone lacking people to speak to and confide my genuine feelings"

EN: So I lived **my** life alone **without** anyone that I could really talk **to**



# Alignment study (Chapters 2 and 7)

- Q1: Why is the alignment so low?

- Idiomatic expressions
- Cross linguistic differences in expression
- Differing word choices & stylistic differences

벌떡 일어섰다  
suddenly stood-up

“[I] stood up suddenly”

EN: I jumped **to** my feet

나는 이렇게 해서 진심을 털어 놓고 이야기 할 사람도 없이 혼자 살아왔다  
I-NOM so (it)was heart-ACC confide speak do person-FOC lacking alone lived

“So it was that I lived alone lacking people to speak to and confide my genuine feelings”

EN: So I lived **my** life alone **without** anyone that I could really talk **to**

# Alignment study (Chapters 2 and 7)

- Q1: Why is the alignment so low?

- Idiomatic expressions
- Cross linguistic differences in expression
- Differing word choices & stylistic differences

벌떡 일어섰다  
suddenly stood-up

“[I] stood up suddenly”

EN: I jumped **to** my feet

나는 이렇게 해서 진심을 털어 놓고 이야기 할 사람도 없이 혼자 살아왔다  
I-NOM so (it)was heart-ACC confide speak do person-FOC lacking alone lived

“So it was that I lived alone lacking people to speak to and confide my genuine feelings”

EN: So I lived **my** life alone **without** anyone that I could really talk **to**

# Alignment study (Chapters 2 and 7)

— — —

- *Q1: Why is the alignment so low?*
- *Q2: Among those aligned, why is the scene agreement so low?*

# Alignment study (Chapters 2 and 7)

— — —

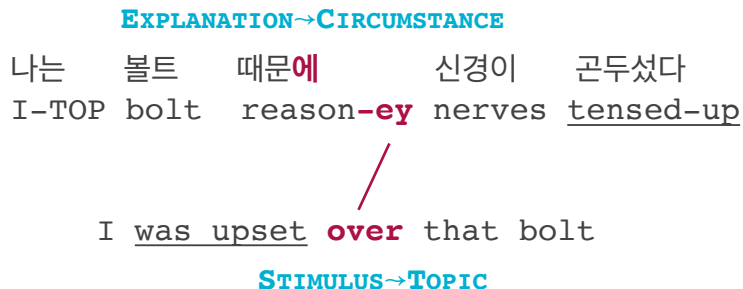
- *Q1: Why is the alignment so low?*
- *Q2: Among those aligned, why is the scene agreement so low?*

While, overall, Korean and English pairs do express parallel meaning via adpositions, the subject matter is handled in ways that the supersenses can't generalize

# Alignment study (Chapters 2 and 7)

- Q1: *Why is the alignment so low?*
- Q2: *Among those aligned, why is the scene agreement so low?*

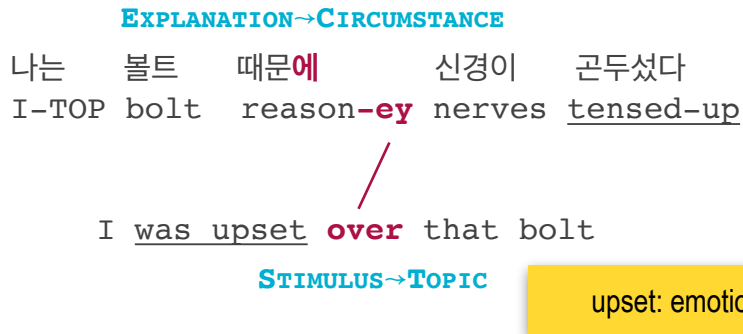
While, overall, Korean and English pairs do express parallel meaning via adpositions, the subject matter is handled in ways that the supersenses can't generalize



# Alignment study (Chapters 2 and 7)

- Q1: *Why is the alignment so low?*
- Q2: *Among those aligned, why is the scene agreement so low?*

While, overall, Korean and English pairs do express parallel meaning via adpositions, the subject matter is handled in ways that the supersenses can't generalize



# Alignment study (Chapters 2 and 7)

- Q1: *Why is the alignment so low?*
- Q2: *Among those aligned, why is the scene agreement so low?*

While, overall, Korean and English pairs do express parallel meaning via adpositions, the subject matter is handled in ways that the supersenses can't generalize

EXPLANATION → CIRCUMSTANCE

나는 볼트 때문에 신경이 곤두섰다  
I-TOP bolt reason-ey nerves tensed-up

tenseness: result of the upset

I was upset over that bolt

STIMULUS → TOPIC

upset: emotion

# Alignment study (Chapters 2 and 7)

— — —

- *Q1: Why is the alignment so low?*
- *Q2: Among those aligned, why so low scene agreement?*
- *Take-away?*



# Alignment study (Chapters 2 and 7)

— — —

- *Q1: Why is the alignment so low?*
- *Q2: Among those aligned, why so low scene agreement?*
- *Take-away?*
  - Natural variations based on linguistic choice

# Alignment study (Chapters 2 and 7)

— — —

- *Q1: Why is the alignment so low?*
- *Q2: Among those aligned, why so low scene agreement?*
- *Take-away?*
  - Natural variations based on linguistic choice
  - Complicated by the fact that EN and KO are both translations of FR

# Alignment study (Chapters 2 and 7)

— — —

- *Q1: Why is the alignment so low?*
- *Q2: Among those aligned, why so low scene agreement?*
- *Take-away?*
  - Natural variations based on linguistic choice
  - Complicated by the fact that EN and KO are both translations of FR
  - Calls for further investigation into nuanced semantics like causality or force dynamics that would aid generalizations

# K-SNACS

## ● 54 coarse-grained supersenses

- Semantic role-like labels
- 4 branches: 3 semantic & 1 pragmatic

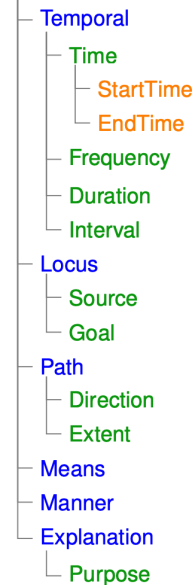
## ● Annotated Corpus

- *The Little Prince*
- 4K postpositional targets & 29 unique p types

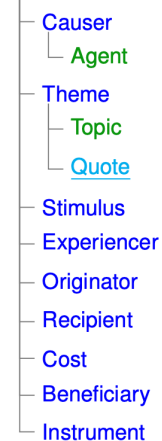
## ● Guidelines and data available at

- [github.com/jdch00/k-snacs](https://github.com/jdch00/k-snacs)

### Circumstance



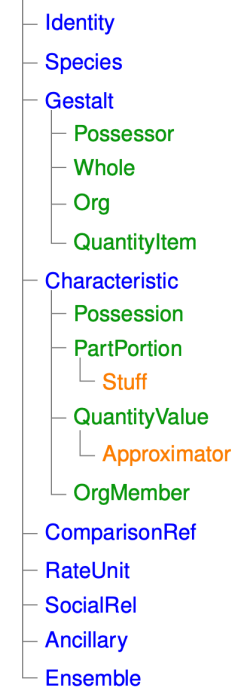
### Participant



### Context



### Configuration



# K-SNACS

## ● 54 coarse-grained supersenses

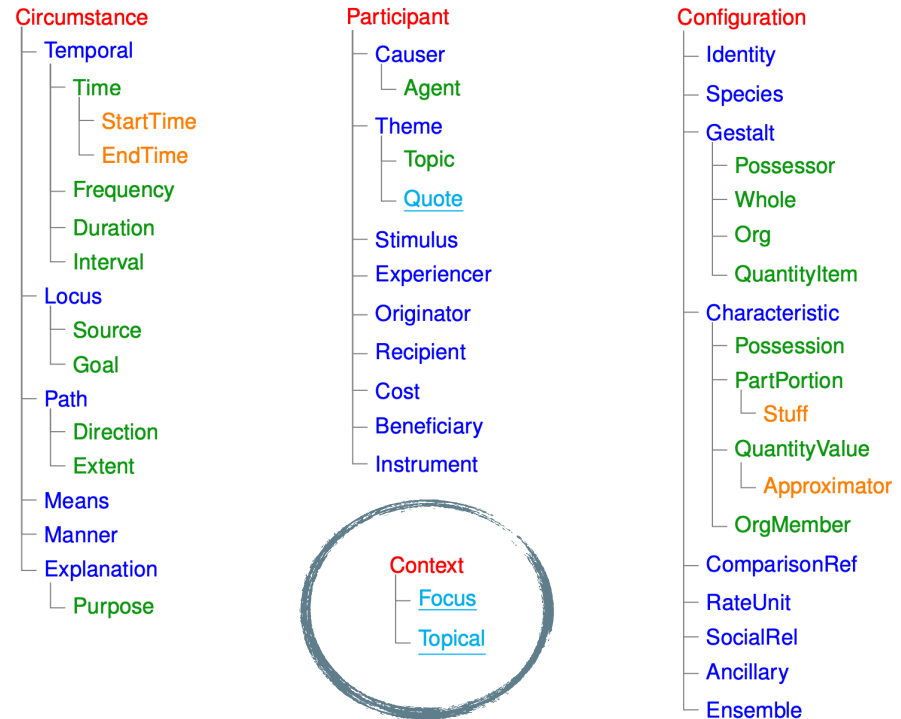
- Semantic role-like labels
- 4 branches: 3 semantic & 1 pragmatic

## ● Annotated Corpus

- *The Little Prince*
- 4K postpositional targets & 29 unique p types

## ● Guidelines and data available at

- [github.com/jdch00/k-snacs](https://github.com/jdch00/k-snacs)



# Conclusion

— — —

- The first annotated corpus of Korean postposition supersenses
- Corpus-based demonstration of adaptability of English SNACS to typologically-distant Korean
- We find that:
  - Korean and English adpositions cover similar semantic grounds
  - Navigate the ground in diverging ways (e.g., Quotatives and Pragmatic postpositions)

# Ongoing & Future Work

— — —

- Investigate the extent in which the newly created labels are applicable to other languages.
- Look further into semantics like causality and force dynamics within the SNACS framework
- Conduct multi-linguistic comparison with other languages adapting SNACS (e.g., Chinese, German, Hindi)

Thank you! (감사합니다!)

Many thanks to:

Vivek Srikumar  
Austin Blodgett

For the many hours of insightful discussions!

This research was supported in part by NSF award IIS-1812778 and grant 2016375 from the United States–Israel Binational Science Foundation (BSF), Jerusalem, Israel.