

# Acquiring the Korean Causatives

**Sun Jae Lee and Jordan Kodner**  
**University of Pennsylvania**

**CLS 55, Chicago**  
**May 16, 2019**

# Outline

- **The Korean Causative Alternation**
- **Productivity Learning**
- **Korean Corpus**
- **Applying the Model**
- **Discussion**

# Background

# The Causative Alternation

- The causative alternation differentiates unaccusative verbs from their corresponding transitives
- Subject of the unaccusative becomes object of the transitive

## Unaccusative

The door opened.

The ice melted.

The ball rolled.

## Transitive

I opened the door.

The sun melted the ice.

Sam rolled the ball.

# The Korean Causative Alternation

- Korean has two causative constructions:
  1. **-히 (-hi)** synthetic, applies to a fixed set of about 40 verbs
    - has phonologically conditioned allomorphs
    - eg, **-이 (-i)**, **-리 (-li)**, **-기 (-gi)**, **-우 (-u)**
    - performs the causative alternation

## Intransitive

철수가 앉는다      ‘Chul-Soo sits’  
sit’

*Chul-Soo-ga anj-neun-da*

철수가 산다      ‘Chul-Soo lives’

*Chul-Soo-ga san-da*

## Transitive

철수를 앉히다 ‘make Chul-Soo

*Chul-Soo-reul anj-hi-da*

철수를 살리다 ‘save Chul-Soo’

*Chul-Soo-reul sal-li-da*

# The Korean Causative Alternation

2. 게 (*ke*) periphrastic, applies to an open class
  - *ke* can make any intransitive verb into a transitive, not limited to unaccusatives

## Intransitive

철수가 먹는다  
eat'

*Chul-Soo-ga meok-neun-da*

철수가 눕는다 'Chul-Soo lies'

*Chul-Soo-ga nup-neun-da*

## Transitive

'Chul-Soo eats' 철수를 먹게하다 'make Chul-Soo eat'

*Chul-Soo-reul meok ke-ha-da*

철수를 눕게하다 'make Chul-Soo lie'

*Chul-Soo-reul nup ke-ha-da*

# Acquiring the Causative Alternation

- When English learners make errors,
- They are characterized by “**over-application**” of alternation<sup>1</sup>
- Causative alternation over-applied because it is productive

## Over-Application of Alternation

### Intransitive

The toy falls.

### Transitive

Adam **fall** toy.<sup>2</sup>

<sup>1</sup> Bowerman 1983, Bowerman & Croft 2008, <sup>2</sup> Adam, Brown Corpus

# Acquiring the Causative Alternation

Prior research on Korean causative acquisition has shown:<sup>1</sup>

1. Errors involve **unexpected non-use** of *-hi*
2. *ke* is productive, whereas *-hi* is not
3. *ke* is acquired by children later than *-hi* is

<sup>1</sup>Choi 1998

# Acquiring the Causative Alternation

- When Korean learners make errors,
- They are characterized by **unexpected non-use of -hi**

## Example of Unexpected Non-use of *-hi* (Yun in Ryu corpus)

초식공룡이 죽으니까 그 공룡이

‘The carnivore dies the herbivore’

Attested: 죽으니까 *juk-eu-nikka* ‘die’

Expected: 죽이니까 *juk-i-nikka* ‘kill’

# A Learning Model

# The Tolerance Principle<sup>1</sup>

- A model for the acquisition of linguistic generalization

<sup>1</sup>Yang 2005, 2016, <sup>2</sup>Chomsky 1955, 1965, Chomsky & Halle 1968, <sup>3</sup>Anderson 1969, *inter alia*, <sup>4</sup>Murray & Forster 2004, <sup>5</sup>Schuler et al 2017

# The Tolerance Principle<sup>1</sup>

- A model for the acquisition of linguistic generalization
- An **evaluation metric**<sup>2</sup> over linguistic hypotheses
  - an **Elsewhere Condition** for ‘rules’ and ‘exceptions’<sup>3</sup>
  - Lexical access is correlated with **frequency-rank**<sup>4</sup>
  - Generally **Zipfian** input distributions

<sup>1</sup>Yang 2005, 2016, <sup>2</sup>Chomsky 1955, 1965, Chomsky & Halle 1968, <sup>3</sup>Anderson 1969, *inter alia*, <sup>4</sup>Murray & Forster 2004, <sup>5</sup>Schuler et al 2017

# The Tolerance Principle<sup>1</sup>

- A model for the acquisition of linguistic generalization
- An **evaluation metric**<sup>2</sup> over linguistic hypotheses
  - an **Elsewhere Condition** for ‘rules’ and ‘exceptions’<sup>3</sup>
  - Lexical access is correlated with **frequency-rank**<sup>4</sup>
  - Generally **Zipfian** input distributions
- **Successfully applied to a wide range of problems**
  - Modern English strong verbs, German noun plurals, Russian and Polish genitives
  - English diatones, American sociolinguistic variables
  - English and Mandarin numeracy, etc.
- **And psychological backing from artificial language learning experiments**<sup>5</sup>

<sup>1</sup>Yang 2005, 2016, <sup>2</sup>Chomsky 1955, 1965, Chomsky & Halle 1968, <sup>3</sup>Anderson 1969, *inter alia*, <sup>4</sup>Murray & Forster 2004, <sup>5</sup>Schuler et al 2017

# Tolerance Principle and Representation

- Forms can be associated with generalizations governing their derivations or memorized as form-derivation pairs
- **Generalization = productive**; **memorization = non-productive**
- So learning a generalization is tantamount to hypothesizing productivity

**Productive generalizations will be extended to unseen forms**

# The Sufficiency Principle<sup>1</sup>

- Reframing of the Tolerance Principle
- Asks whether the learner has received enough evidence for a generalization
- Given a hypothesized generalization  $R$  operating over a class  $C$ , quantitatively define the number of **(yet) unattested forms** below which the generalization is tenable

# The Sufficiency Principle<sup>1</sup>

- Reframing of the Tolerance Principle
- Asks whether the learner has received enough evidence for a generalization
- Given a hypothesized generalization  $R$  operating over a class  $C$ , quantitatively define the number of **(yet) unattested forms** below which the generalization is tenable

$N = |C|$  by types

$M = |\text{types} \in C \text{ attested obeying } R|$

$\theta = \text{threshold} = N / \ln N$

Evidence is **sufficient** if

$$N - M < N / \ln N$$

# $N$ and $M$ Vary over Individual Development

- $N$  is the number of class members a child has learned *so far*
- $N$  and  $M$  grow as the learner's vocabulary grows
- Children fall into and out of productivity during development

## How the SP applies to the acq of *-hi* and *ke*

- The child does not know *a priori* which of the constructions are productive

## How the SP applies to the acq of *-hi* and *ke*

- The child does not know *a priori* which of the constructions are productive

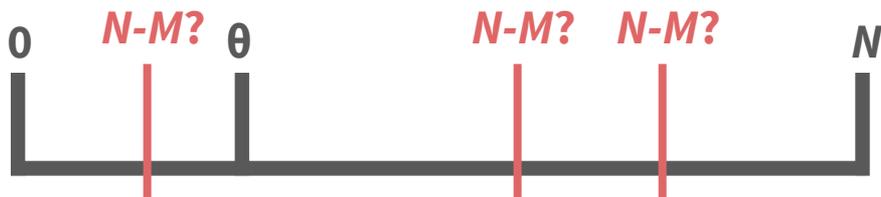
Given a potential semantic generalization (e.g., unacc~trans alternator) that can be associated with *-hi* or *ke* in the input,

- Are there enough instances of that construction applying to those verbs that I can assume I can apply it to similar verbs?
- If so, apply it productively to those obeying the generalization
- If not, assume it is lexical and memorize word-by-word

## E.g., Visualization for Unaccusatives and *-hi*

$N$  = # of unaccusative verbs

$M$  = # attested with *-hi*



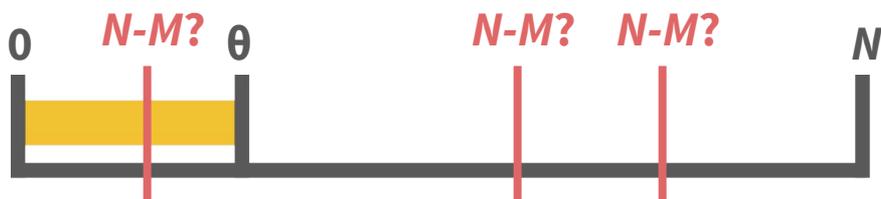
$N-M$  = # of unaccusative verbs not (yet) attested with *-hi*

$\theta$  = threshold =  $N / \ln N$

## E.g., Visualization for Unaccusatives and *-hi*

$N$  = # of unaccusative verbs

$M$  = # attested with *-hi*



$N-M$  = # of unaccusative verbs not (yet) attested with *-hi*

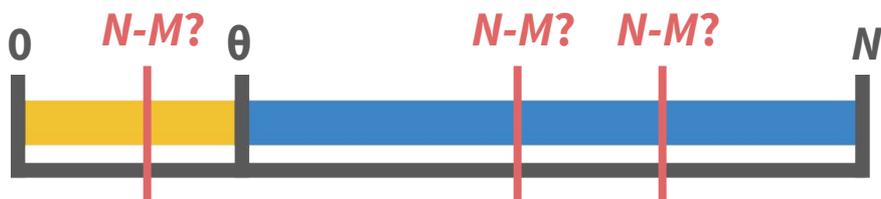
$\theta$  = threshold =  $N / \ln N$

- If  $N-M$  is below  $\theta$ , enough *-hi* unaccusatives are attested to render *-hi* productive

## E.g., Visualization for Unaccusatives and *-hi*

$N$  = # of unaccusative verbs

$M$  = # attested with *-hi*



$N-M$  = # of unaccusative verbs not (yet) attested with *-hi*

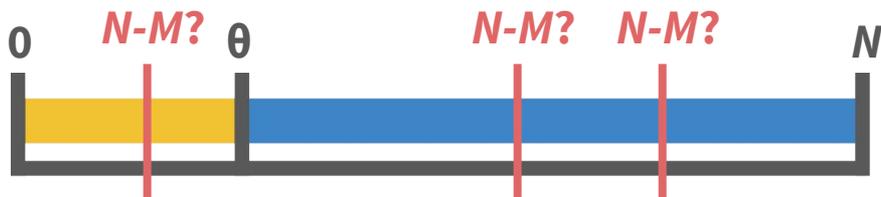
$\theta$  = threshold =  $N / \ln N$

- If  $N-M$  is below  $\theta$ , enough *-hi* unaccusatives are attested to render *-hi* productive
- If  $N-M$  is above  $\theta$ , memorize the individual *-hi* unaccusatives

## E.g., Visualization for Unaccusatives and *-hi*

$N$  = # of unaccusative verbs

$M$  = # attested with *-hi*



$N-M$  = # of unaccusative verbs not (yet) attested with *-hi*

$\theta$  = threshold =  $N / \ln N$

- If  $N-M$  is below  $\theta$ , enough *-hi* unaccusatives are attested to render *-hi* productive
- If  $N-M$  is above  $\theta$ , memorize the individual *-hi* unaccusatives
- The judgment may change based on new evidence
- Equivalent calculations for *ke* and other semantic generalizations

# Korean Data

# Korean Child-Produced and -Directed Speech

- Yun's child-directed (CDS) and child-produced (CPS) speech in CHILDES Ryu
- Child-produced causative utterances were catalogued
  - Divided these into “adult-like” and “error” productions
- All CDS verbs were sorted into unaccusative/unergative/stative and available causative formations were identified
- Statives are common in Korean but rare in English
  - 뜨겁다 *tteu-geop-da* ‘be hot’ → 뜨겁게 하다 *tteu-geop ke-ha-da* ‘make hot’
  - 조용하다 *jo-yong-ha-da* ‘be quiet’ → 조용하게 하다 *jo-yong ha-ke-ha-da* ‘make quiet’

(Merging statives with unaccusatives or unergatives does not change outcomes)

## Yun's Learner “Errors”

- Predominantly unexpected non-use of *-hi*
- Unexpected use of *ke* are instances where *-hi* would have been preferred

Korean Error (Yun)	Count
<i>-hi</i> unexpected use	1
<i>-hi</i> unexpected non-use	6
<i>ke</i> unexpected use	2
<i>ke</i> unexpected non-use	0
Total CDS utterances	81,577
Total CPS utterances	38,356

# Accounting for Acquisition

# Calculations

- Use Sufficiency Principle to calculate productivity of Korean causative constructions **both for early learners and adults**
- Modeled early learner's input using Yun CDS
  - CDS is often used to approximate child linguistic experience<sup>1</sup>
  - CDS models items in the child lexicon and the proportion attested with each causative type
- Modeled adult knowledge of the Yun CDS verbs by classifying according to native speaker judgments

<sup>1</sup>Nagy & Anderson 1984, Yang 2016, etc.

# Modeling an Early Learner's Productivity Judgments

- For an early learner, neither construction is productive - they are both lexical
- Expect under-application because there is no way to extend either construction to verbs not yet learned
- More *-hi* verbs are attested than *ke* verbs

In Yun CDS	<i>N</i>	$\theta$	<i>M -hi</i>	<i>-hi</i> Productive?	<i>M ke</i>	<i>ke</i> Productive?
Unaccusatives	25	7.6	12	<i>N-Mhi</i> =13, <b>no</b>	4	<i>N-Mke</i> =21, <b>no</b>
Unergatives	129	26.5	12	<i>N-Mhi</i> =117, <b>no</b>	3	<i>N-Mke</i> =126, <b>no</b>
Statives	74	17.2	1	<i>N-Mhi</i> =73, <b>no</b>	6	<i>N-Mke</i> =68, <b>no</b>

# Modeling an Adult's Productivity Judgments

- For an adult, *-hi* is not productive for any class - it is still lexical
- *ke* is productive for all verbs
- At some point during development, learners must hear enough verb types with *ke* causatives for it to become productive

Adult Judgment	<i>N</i>	$\theta$	<i>M -hi</i>	<i>-hi</i> Productive?	<i>M ke</i>	<i>ke</i> Productive?
Unaccusatives	25	7.6	16	<i>N-Mhi</i> =9, <b>no</b>	25	<i>N-Mke</i> =0, <b>YES</b>
Unergatives	129	26.5	11	<i>N-Mhi</i> =118, <b>no</b>	128	<i>N-Mke</i> =1, <b>YES</b>
Statives	74	17.2	3	<i>N-Mhi</i> =71, <b>no</b>	66	<i>N-Mke</i> =8, <b>YES</b>

# Discussion

# Accounting for Korean Acquisition Observations

## Unexpected Non-use of *-hi*

- SP defines it as non-productive (lexical-only) for young learners and adults

## Only *ke* is productive

- SP is consistent with this for adults

## *ke* is acquired later than *-hi*

- For early learners, both are unproductive, but more types are attested with *-hi*
- Since both are first learned word-by-word (ie, non-productively), children can use *-hi* with more verbs than *ke*
- Gives the appearance of later acquisition

# End

## Acknowledgements:

- Charles Yang
- Ava Irani

## Support:

NDSEG (2nd author)

## Example of Non-Use of *-hi* (Ryu Yun corpus)

손이 올라가다가 다쳤어

‘My hand was risen and got hurt.’

Attested: 올라가다 *ol-la-gada* ‘go up’

Expected: 올리다 *ol-li-da* ‘raise’

## Example of Unexpected Use of -ke (Ryu Yun corpus)

뜨겁게 해야 돼

‘You have to make it hot.’

Attested:      뜨겁게 하다      *tteu-geop-ke-hada* ‘make hot’

Expected:      데우다      *de-u-da* ‘heat’

# The Korean Causative Alternation

- Yun and Ross corpora are comparable in size (both CDS and CPS)
- Show contrast between English and Korean in number of each error type
- While English learners show **over-application** of the alternation, Korean learners show **unexpected non-use of *-hi***

Korean Error (Yun)	Count	English Error (Ross)	Count
<i>-hi</i> unexpected use	1	Over-application	10
<i>-hi</i> unexpected non-use	6	Under-application	0
<i>ke</i> unexpected use	2		
<i>ke</i> unexpected non-use	0		
Total CDS utterances	81,577	Total CDS utterances	82,466
Total CPS utterances	38,356	Total CPS utterances	35,912