Born To Be Gradient

Predicting Exceptions of Compound Tensing in Korean

Hyunjung Lee

OCP 2019, University of Verona

17 January 2019

UNIVERSITÄT LEIPZIG

INTERACTION OF GRAMMATICAL BUILDING BLOCKS

Nutshell

How to deal with exceptionality ?

- **Compound Tensing (CT)** in Korean unexpectedly fails to apply to certain Noun-Noun compounds (Jun 2001; Zuraw 2011; Ito 2014; Kim 2016).
- Should this exceptionality be dealt with the grammar or through lexicalization ?

Gradient Symbolic Representation

- I argue for an account in terms of **Gradient Symbolic Representations** (**GSR** ; Smolensky and Goldrick, 2016, Rosen 2016).
- The intrinsic property of GSR captures **the nature of gradient inclination for CT**, which is impossible with other systems.

Learnability

- An **error-driven algorithm** also shows tat the scalar activities are learnable.

Data

Laryngeal contrasts

- Korean has a three-way distinction in terms of laryngeal contrast in obstruents

(1)

(a)	/pul/	→ [**p**ul]	'fire'
(b)	/p^{h}ul/	→ [**p^h**ul]	'grass'
(c)	/p'ul/	→ [**p'**ul]	'horn'

Laryngeal contrasts

- Korean has a three-way distinction in terms of laryngeal contrast in obstruents

(1)

(a)	/pul/	→ [**p**ul]	'fire'
(b)	/p^{h}ul/	→ [**p^{h}**ul]	'grass'
(c)	/p'ul/	→ [**p'**ul]	'horn'

Compound Tensing

- **Compound Tensing (CT)** :
  When a **compound** consist of two nouns, W_A and W_B, initial plain obstruents of W_Bs undergo junctural processes including **obstruent tensification**.

(2)

(a)	/hɛ/ + /pic/	→ [hɛ.**p'**it]	post Vowel
(b)	/kailɨl/ + /pi/	→ [ka.ɨl.**p'**i]	post Lateral
(c)	/pom/ + /pi/	→ [pom.**p'**i]	post Nasal
(d)	/pok/ + /pi/	→ [pok.**p'**i]	post obstruent

Exceptionality

- 23% noun-noun compounds exceptioanlly does not undergo CT in a random fashion
(Jun 2015; Zuraw 2011; Ito 2014; Kim 2016).

(3)

	Regular Pattern		Exception		
(a)	/hɛ/ + /pap/	→ [hɛ.**p’**ap]	(e)	/koŋ/ + /pap/	→ [koŋ.**p**ap]
(b)	/hɛ/ + /kuks’u/	→ [hɛ.**k’**uks’u]	(f)	/koŋ/ + /kuks’u/	→ [koŋ.**k**uk.s’u]
(c)	/pipim/ + /pap/	→ [pi.pim.**p’**ap]	(g)	/pipim/ + /kuks’u/	→ [pi.pim.**k**uk.s’u]
(d)	/koŋ/ + /karu/	→ [koŋ.**k’**a.ru]	(h)	/hɛ/ + /toci/	→ [hɛ.**t**o.ci]

Exceptionality

- 23% noun-noun compounds exceptioanlly does not undergo CT in a random fashion
(Jun 2015 ; Zuraw 2011 ; Ito 2014 ; Kim 2016).

(3)

	Regular Pattern			Exception	
(a)	/hɛ/ + /pap/	→ [hɛ.**p'**ap]	(e)	/koŋ/ + /pap/	→ [koŋ.**p**ap]
(b)	/hɛ/ + /kuks'u/	→ [hɛ.**k'**uks'u]	(f)	/koŋ/ + /kuks'u/	→ [koŋ.**k**uk.s'u]
(c)	/pipim/ + /pap/	→ [pi.pim.**p'**ap]	(g)	/pipim/ + /kuks'u/	→ [pi.pim.**k**uk.s'u]
(d)	/koŋ/ + /karu/	→ [koŋ.**k'**a.ru]	(h)	/hɛ/ + /toci/	→ [hɛ.**t**o.ci]

Gradient Pattern of Tensing

- The compound tensing exhibit continuum of **gradient preferences** depending on **both the conjuncts $\mathbf{W}^A$, $\mathbf{W}^B$** in the compound.

(4)

(a)	/hɛ/	+ /pap/	→ [hɛ.**p'**ap]
(b)	/hɛ/	+ /kuks'u/	→ [hɛ.**k'**uks'u]
(c)	/hɛ/	+ /kali/	→ [hɛ.**ka**.li]
(d)	/pipim	+ /pap/	→ [pi.pim.**p'**ap]
(e)	/pipim	+ /kuks'u/	→ [pi.pim.**kuk**.s'u]
(f)	/pipim	+ /kali/	→ [pi.pim.**ka**.li]
(g)	/koŋ/	+ /pap/	→ [koŋ.**pap**]
(h)	/koŋ/	+ /kuks'u/	→ [koŋ.**kuk**.s'u]
(i)	/koŋ/	+ /kali/	→ [koŋ.**ka**.li]

Gradient Pattern of Tensing

- The compound tensing exhibit continuum of **gradient preferences** depending on **both the conjuncts** $\mathbf{W}^A$, $\mathbf{W}^B$ in the compound.

(4)

(a)	/hɛ/	+ /pap/	→ [hɛ.**p’**ap]
(b)	/hɛ/	+ /kuks’u/	→ [hɛ.**k’**uks’u]
(c)	/hɛ/	+ /kali/	→ [hɛ.**k**a.li]
(d)	/pipim	+ /pap/	→ [pi.pim.**p’**ap]
(e)	/pipim	+ /kuks’u/	→ [pi.pim.**k**uk.s’u]
(f)	/pipim	+ /kali/	→ [pi.pim.**k**a.li]
(g)	/koŋ/	+ /pap/	→ [koŋ.**p**ap]
(h)	/koŋ/	+ /kuks’u/	→ [koŋ.**k**uk.s’u]
(i)	/koŋ/	+ /kali/	→ [koŋ.**k**a.li]

Gradient Pattern of Tensing

- The compound tensing exhibit continuum of **gradient preferences** depending on **both the conjuncts** $\mathbf{W}^A$, $\mathbf{W}^B$ in the compound.

(4)

(a)	/hɛ/	+ /pap/	→ [hɛ.**p'**ap]
(b)	/hɛ/	+ /kuks'u/	→ [hɛ.**k'**uks'u]
(c)	/hɛ/	+ /kali/	→ [hɛ.**k**a.li]
(d)	/pipim	+ /pap/	→ [pi.pim.**p'**ap]
(e)	/pipim	+ /kuks'u/	→ [pi.pim.**k**uk.s'u]
(f)	/pipim	+ /kali/	→ [pi.pim.**k**a.li]
(g)	/koŋ/	+ /pap/	→ [koŋ.**p**ap]
(h)	/koŋ/	+ /kuks'u/	→ [koŋ.**k**uk.s'u]
(i)	/koŋ/	+ /kali/	→ [koŋ.**k**a.li]

Gradient Pattern of Tensing

- The compound tensing exhibit continuum of **gradient preferences** depending on **both the conjuncts $\mathbf{W}^A$, $\mathbf{W}^B$** in the compound.

(4)

(a)	/hɛ/	+ /pap/	→ [hɛ.**p’**ap]
(b)	/hɛ/	+ /kuks’u/	→ [hɛ.**k’**uks’u]
(c)	/hɛ/	+ /kali/	→ [hɛ.**k**a.li]
(d)	/pipim	+ /pap/	→ [pi.pim.**p’**ap]
(e)	/pipim	+ /kuks’u/	→ [pi.pim.**k**uk.s’u]
(f)	/pipim	+ /kali/	→ [pi.pim.**k**a.li]
(g)	/koŋ/	+ /pap/	→ [koŋ.**p**ap]
(h)	/koŋ/	+ /kuks’u/	→ [koŋ.**k**uk.s’u]
(i)	/koŋ/	+ /kali/	→ [koŋ.**k**a.li]

Gradient Pattern of Tensing

- The compound tensing exhibit continuum of **gradient preferences** depending on **both the conjuncts** $\mathbf{W}^A$, $\mathbf{W}^B$ in the compound.

(5)

(a)	/hɛ/	+ /pap/	→ [hɛ.**p’**ap]
(b)	/pipim/	+ /pap/	→ [pi.pim.**p’**ap]
(c)	/koŋ/	+ /pap/	→ [koŋ.**p**ap]
(d)	/hɛ/	+ /kuks’u/	→ [hɛ.**k**uks’u]
(e)	/pipim/	+ /kuks’u/	→ [pi.pim.**k’**uk.s’u]
(f)	/koŋ/	+ /kuks’u/	→ [koŋ.**k**uk.s’u]
(g)	/hɛ/	+ /kali/	→ [hɛ.**k**a.li]
(h)	/pipim/	+ /kali/	→ [pi.pim.**k**a.li]
(i)	/koŋ/	+ /kali/	→ [koŋ.**k**a.li]

Gradient Pattern of Tensing

- The compound tensing exhibit continuum of **gradient preferences** depending on **both the conjuncts $\mathbf{W}^A$, $\mathbf{W}^B$** in the compound.

(5)

(a)	/hɛ/	+ /pap/	→ [hɛ.**p'**ap]
(b)	/pipim/	+ /pap/	→ [pi.pim.**p'**ap]
(c)	/koŋ/	+ /pap/	→ [koŋ.**pap**]
(d)	/hɛ/	+ /kuks'u/	→ [hɛ.**k**uks'u]
(e)	/pipim/	+ /kuks'u/	→ [pi.pim.**k'**uk.s'u]
(f)	/koŋ/	+ /kuks'u/	→ [koŋ.**kuk**.s'u]
(g)	/hɛ/	+ /kali/	→ [hɛ.**ka**.li]
(h)	/pipim/	+ /kali/	→ [pi.pim.**ka**.li]
(i)	/koŋ/	+ /kali/	→ [koŋ.**ka**.li]

Gradient Pattern of Tensing

- The compound tensing exhibit continuum of **gradient preferences** depending on **both the conjuncts $\mathbf{W}^A$, $\mathbf{W}^B$** in the compound.

(5)

(a)	/hɛ/	+ /pap/	→ [hɛ.**p’**ap]
(b)	/pipim/	+ /pap/	→ [pi.pim.**p’**ap]
(c)	/koŋ/	+ /pap/	→ [koŋ.**pap**]
(d)	/hɛ/	+ /kuks’u/	→ [hɛ.**k**uks’u]
(e)	/pipim/	+ /kuks’u/	→ [pi.pim.**k’**uk.s’u]
(f)	/koŋ/	+ /kuks’u/	→ [koŋ.**kuk**.s’u]
(g)	/hɛ/	+ /kali/	→ [hɛ.**ka**.li]
(h)	/pipim/	+ /kali/	→ [pi.pim.**k**a.li]
(i)	/koŋ/	+ /kali/	→ [koŋ.**ka**.li]

Gradient Pattern of Tensing

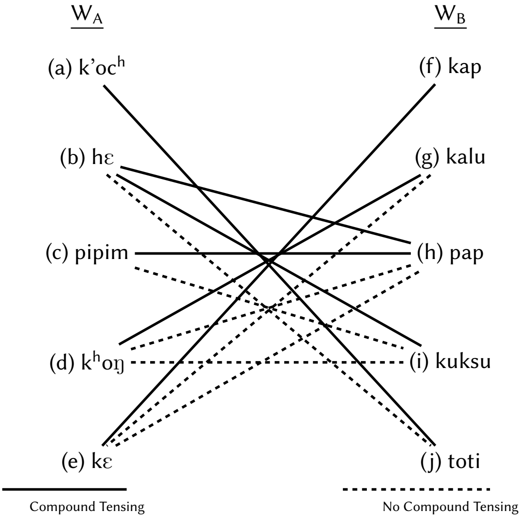
- The compound tensing exhibit continuum of **gradient preferences** depending on **both the conjuncts** $\mathbf{W}^A$, $\mathbf{W}^B$ in the compound.

(5)

(a)	/hɛ/	+ /pap/	→ [hɛ.**p'**ap]
(b)	/pipim/	+ /pap/	→ [pi.pim.**p'**ap]
(c)	/koŋ/	+ /pap/	→ [koŋ.**p**ap]
(d)	/hɛ/	+ /kuks'u/	→ [hɛ.**k**uks'u]
(e)	/pipim/	+ /kuks'u/	→ [pi.pim.**k'**uk.s'u]
(f)	/koŋ/	+ /kuks'u/	→ [koŋ.**k**uk.s'u]
(g)	/hɛ/	+ /kali/	→ [hɛ.**k**a.li]
(h)	/pipim/	+ /kali/	→ [pi.pim.**k**a.li]
(i)	/koŋ/	+ /kali/	→ [koŋ.**k**a.li]

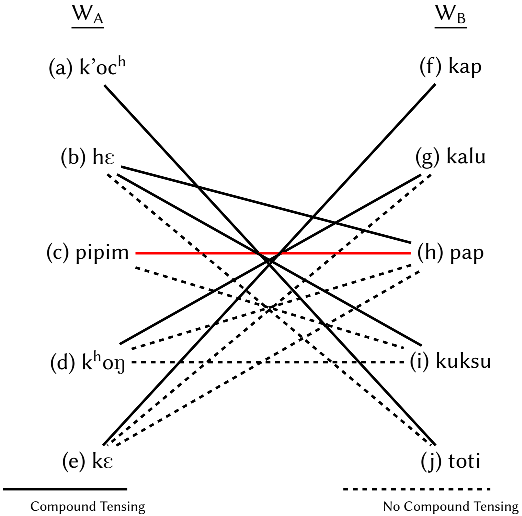
Gradient Pattern of Tensing

(6) Gradient patterns for compounding tensing



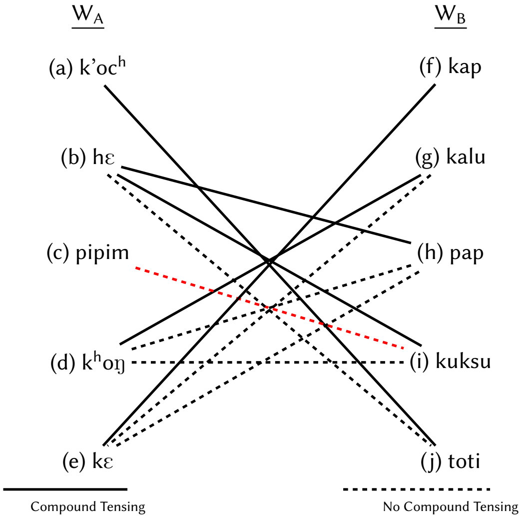
Gradient Pattern of Tensing

(6) Gradient patterns for compounding tensing



Gradient Pattern of Tensing

(6) Gradient patterns for compounding tensing



Gradient Pattern of Tensing

There is no way in standard rule-based (Chomsky and Halle, 1968) or Optimality theory frameworks (Prince and Smolensky, 1993) where features are binary or privative, to give a word a feature that will determine its precise degree of preference for CT .

Proposal

Gradient Symbolic Representation

- Symbols in a linguistic representation can have **different activities** :

  'Symbols are discrete but their degree of presence in a given linguistic representation is continuously gradient' (Smolensky and Goldrick, 2016, 2)

- (Continuous) Numerical strength from 0 to 1 can be associated to input

- Output elements are all fully active (1) as descrete forms

Gradient Symbolic Representation

- The underlying structure is grammatically computed inside **Harmonic Grammar** (Legendre et al. 1990)
- It can predict lexical exceptions :
  - Elements in the underlying representation of a morpheme can be **too weak** to undergo/trigger a certain process
  - Elements associated with different activity can be **strong enough** to undergo/trigger the same process

Claim

- I suggest that each edge of nouns in Korean may have **floating feature [cg]** (Zoll 1996) with **gradient activitity** in the underlying structures (Rosen 2016, 2018)

(7)

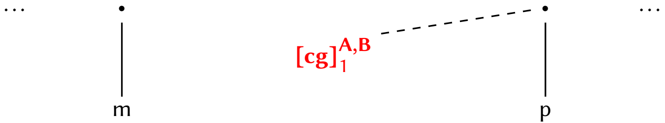
… • • …

m $[\mathrm{cg}]^{A}_{0.4}$ $[\mathrm{cg}]^{B}_{0.2}$ k

Claim

- CT occurs by the **coalescence** of two stem-specific, partially activated floating [cg] features and **docking** to the root node

(8)



Claim

- Only when the additive combination of these features $[cg]^{A,B}$exceeds some threshold Σ does tensing occur.

(9) A hierarchy of 5-level of activation values for compounding tensing

$[cg]_A$ / $[cg]_B$	0 ————————————-1				
0	✗	✗	✗	✗	✓
	✗	✗	✗	✓	✓
	✗	✗	✓	✓	✓
	✗	✓	✓	✓	✓
1	✓	✓	✓	✓	✓

Claim

- Only when the additive combination of these features $[\text{cg}]^{A,B}$ exceeds some threshold Σ does tensing occur.

(9) A hierarchy of 5-level of activation values for compounding tensing

$[\text{cg}]_A$ / $[\text{cg}]_B$	0 —————————————-1				
0	✗	✗	✗	✗	✓
	✗	✗	✗	✓	✓
	✗	✗	✓	✓	✓
	✗	✓	✓	✓	✓
1	✓	✓	✓	✓	✓

Constraints

- MAX[cg] : Input must have output correspondents.
  It **rewards** underlying activity that makes it to the surface.

→ i.e., the more strength the feature bears, the more rewards it induces when it realizes

- IDENT[cg] : The specification for the feature [cg] of an input segment must be preserved in its output correspondent.

→ i.e., It **penalizes** the feature change

- UNIFORMITY[cg] : No feature [cg] in the output has multiple correspondents in the input.

→ i.e., 'No coalescence'

Optimization

- This analysis accounts for the gradient nature of CT.

- The Harmony of the representation τ is :

(10) $$H(r) = 1 \cdot \mathbb{C}_{\text{Max[cg]}}(r) - 0.6 \cdot \mathbb{C}_{\text{Ident[cg]}}(r) - 0.1 \cdot \mathbb{C}_{\text{Uniformity[cg]}}(r)$$

- The candidate with maximal harmony in its candidate set is the optimal output

Optimization : Compound Tensing

W_A : /pipim/ - τ : 0.4, W_B : /pap/ - τ : 0.4

(11) *T_1. pipim + pap → [pi.pim.**p'**ap]*

… • m $[cg]^{x}_{0.4}$ $[cg]^{y}_{0.4}$ • p …	MAX ([c.g]) $w = 100$	IDENT ([c.g]) $w = -60$	UNIFORMITY ([c.g]) $w = -10$	H
O_1 : … • m • p …				0
☞ O_2 : … • m $[cg]^{x,y}_{1}$ • p …	(0.4+0.4)	1	1	10

- The sum of additive feature [cg] from two conjuncts are **strong enougth** to undergo CT

Optimization : No Compound Tensing

W_A : /pipim/ - τ : 0.4, W_B : /kuksu/ - τ : 0.2

(12) *T_2. pipim + kuksu → [pi.pim.**k**uk.s'u]*

… m $[cg]^x_{0.4}$ $[cg]^y_{0.2}$ k …	MAX ([c.g]) $w = 100$	IDENT ([c.g]) $w = -60$	UNIFORMITY ([c.g]) $w = -10$	H
☞ O_1 : … m k …				0
O_2 : … m $[cg]^{x,y}_1$ k …	(0.4+0.2)	1	1	−10

- The total sum of the feature [cg] of 'pipim' and 'kuksu' is **too weak** to undergo tensification.

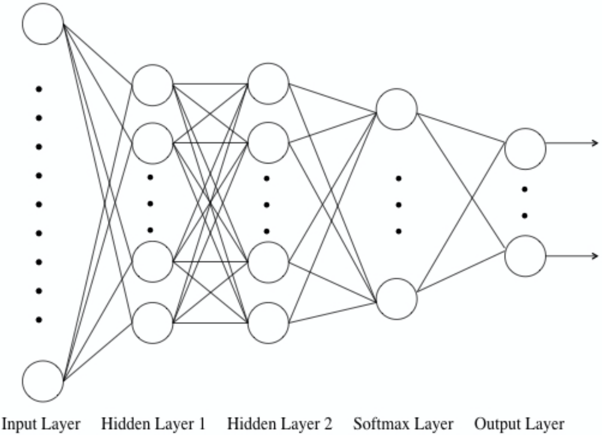
Why Gradience?

Not only do words that occur as **the second conjunct** of a compound exhibit gradient preferences for [cg], but **the first conjunct** in the compound also arguably exhibits the same kind of **gradient preference for triggering tensing** in the word that follows it.

Learnability

The error-driven learning algorithm

(13) An Architecture of Convolutional Neural Network



The error-driven learning algorithm

Step 1 : Initialization

1. A learning algorithm was trained through Convolutional Neural Network (Mikolov et al. 2013)
   - It consists of 2 hidden and 1 softmax layers
2. Activation levels for [cg] of the W^A s and W^Bs were initialized at 0.5
3. Constraints MAX and IDENT were initialized with unit values
4. UNIFORMITY and LINEARITY have fixed values
5. The threshold levels for the sum values of [cg] for compounds were set at 0.7

The error-driven learning algorithm

Step 2 : Iteration

1. The compounds [W^A+ W^B] are **evaluated** on each iteration to check whether each gross effect of CT is correctly derived;
   - will get a **reward** +10 if the correct pattern is derived,
   - will get a **penalty** -5 if the wrong pattern is derived

2. When two coalescing activations [cg] require **adjusting**,
   - It randomly refills the both values of [cg] by either decrementing or incrementing them (a stepsize of 0.05)
   - MAX and IDENT adjust their weights slightly adjusted through a simulated-annealing process (De Vicente et al. 2003) [1]

Step 3 : Convergence

- After 16533 iterations (i.e., when the algorithm can predict all the training set data of CT corretly) the training of this learning was converged.

1. with a decaying temperature T and random Gaussian noise N with m = 0 and $s.d.$= 0.05

Results

Results	
Average of iterations	32
Final Value of MAX	1.121
Final Value of IDENT	0.69
The number of activation levels for W^A	5
The number of activation levels for W^B	5

Conclusion

Conclusion

1. This **GSR analysis** can predict all the patterns of exceptional non-undergoer of Compound Tensing successfully without any redundancy rules
2. The intrinsic property of GSR enables the elements to bear a scalar strength and to **capture the lexical exception of alternation** in the same context
3. Although the distinction is not visible on the surface, there are reasons to believe that obstruents in Korean has diverse patterns of different underlying structures with a **gradiently active feature [cg]**
4. The learning algorithm also supports that this scaler grammar is **learnable**

Contact Information

Hyunjung Lee
Univeristy of Leipzig
hyunjung.lee@uni-leipzig.de

References I

- De Vicente, Juan, Juan Lanchares, & Roman Hermida (2003) Placement by thermodynamic simulated annealing. *Physics Letters A* 317.5-6 415-423.
- Inkelas, Sharon & Cho, Young-Mee Yu (1994). Post-obstruent tensification in Korean and geminate inalterability.Theoretical issues in Korean linguistics, 45.
- Ito, Chiyuki (2014). Compound tensification and laryngeal co-occurrence restrictions in Yanbian Korean. Phonology 31.3, 349-398.
- Jun, Jongho. (2015) Korean n-insertion : A mismatch between data and learning. Phonology 32(3), 417 - 458.
- Kim, Seoyoung (2016) Phonological trends in Seoul Korean compound tensification. MA thesis, Seoul National University, Korea.
- Mikolov, Tomas, Sutskever, Ilya and Chen, Kai, Corrado, Greg S & Dean, Jeff (2013). Distributed Representations of Words and Phrases and their Compositionality. In *Proceedings of NIPS 2013.*
- Legendre, Geraldine, Yoshiro Miyata & Paul Smolensky (1990). Harmonic grammar – a formal multi-level connectionist theory of linguistic well-formedness : Theoretical foundations. *Proceedings of the 12th annual conference of the cognitive science society*. 388–395.

References II

- Smolensky, Paul & Matthew Goldrick (2016). Gradient symbolic representations in grammar : The case of French Liaison. *ROA 1286*
- Rosen, Eric (2016), Predicting the unpredictable : Capturing the apparent semi-regularity of rendaku voicing in Japanese through harmonic grammar, in E.Clem, V.Dawson, A.Shen, A. H.Skilton, G.Bacon, A.Cheng and E. H.Maier, eds, *Proceedings of BLS 42 Berkeley Linguistic Society*, 235–249.
- Rosen, Eric (2018). Evidence for gradient input features from Sino-Japanese compound accent. poster, presented at AMP 2018, San Diego.
- Zoll, Cheryl (1996). Parsing below the segment in a constraint-based framework, PhD thesis, UC Berkeley.
- Zuraw, Kie. (2011) Predicting sai-siot in Korean compound nouns : Phonological and non-phonological factors. Handout presented at the 21st Japanese/Korean Linguistic Conference, Seoul National University.