Acquiring the Korean Causatives*

Sun Jae Lee and Jordan Kodner
University of Pennsylvania

1 Introduction

The acquisition of argument structure is one of the classic problems in child language acquisition (Baker 1979, Pinker 1984, Landau & Gleitman 1985, Pinker 1989, Bowerman & Croft 2008, Arunachalam 2015). Argument structure alternations present a challenge to learners because they are only applicable to certain verbs. While there are identifiable semantic conditions on verbs' participation in the alternations, there are often apparent exceptions. And when considering the sparsity of early linguistic input in general, it is far from given that every verb that can participate in some alternation will actually be attested in a given child's input. Thus, the learner is tasked with generalizing argument structure alternations to unseen verbs while simultaneously limiting them to only the verbs which support them. This is Baker's Paradox (Baker 1979).

One such argument structure alternation that has received considerable analysis in English has been the *causative alternation*, which describes the pattern by which some unaccusative verbs may appear in transitive constructions in English as in (1). Not all unaccusative verbs actually permit the alternation (2), and it is up to the learner to determine which verbs allow it. Those that disallow it may instead distinguish the unaccusative and causative lexically (*eat*∼*feed*, *rise*∼*raise*) or rely on periphrastic constructions with *make*, or *cause*, among others.

(1) English causative alternation
  a. The ice melted.
     Alice melted the ice.
  b. The door is opening.
     Bob is opening the door.

(2) English non-alternation
  a. The mosquito died.
     Alice *died/killed the frog.
     Alice caused the frog to die.
  b. The ball will fall.
     Bob *will fall/will drop the ball.
     Bob will make the ball fall.

*We would like to thank Charles Yang for his insights and support during the completion of this project, Ava Irani for her counts for Ross, and the reviewers, audience, and editors for CLS 55.

English-learning children over-apply the causative alternation during their development (Pinker 1989, Bowerman & Croft 2008, Irani 2019) by using the same verb form in both intransitive and transitive contexts. This can be seen as an overgeneralization of the alternation in several frameworks.

Although there are similarities in the patterning of causatives across languages, there are also systemic differences (Shibatani & Pardeshi 2002), yet less work has been done on the acquisition of causative patterns outside of English. A cross-linguistic comparison of language acquisition patterns stands to shed light on the relationship between language-independent learning processes and language-specific input on the development of language.

We perform such an analysis for Korean utilizing the *Sufficiency Principle* (SP; Yang 2016; §6.1.3), a recent input-driven framework for argument structure learning, in conjunction with recently available corpora of child-directed and child-produced speech to investigate the developmental trajectory of the causative patterns in Korean. Additionally, we compare and contrast developmental outcomes in Korean with recent work modeling English learners' acquisition of causatives (Irani 2019) and find that the same learning mechanisms, together with language-specific input, account for the different developmental trajectories and learning outcomes in both languages.

The following section describes the Korean causative alternation and its morphological and periphrastic causative patterns. Next, Section 3 presents background on Korean learners' anomalous causative productions and introduces the Sufficiency Principle. In Section 4, the SP is applied to input drawn from Korean CHILDES and to adult grammaticality judgments to predict the productivity of both causative patterns. Finally, Section 5 summarizes this work and discusses its implications.

2 Korean causatives

Korean has two causative constructions which we will call the *-hi causative* and *ge causative*.[1] The former is formed morphologically with an affix *-hi* attached to the root (3b). *-hi* has several allomorphs (*-i*/*-hi*/*-li*/*-gi*/*-wu*/*-gwu*/*-jwu*) which are largely but not entirely phonologically predictable (Yeon 1991), and it can only be applied to a fixed set of native verbs, most of which are unaccusatives (Yeon 1991, Park 1994, Choi 1999, Shibatani & Pardeshi 2002).

(3) Unaccusative alternation with *-hi* (Park 1994 (49a-b))

a. *mul-i* *eol-eot-da*
water-NOM freeze-PAST-IND
'The water froze.'

b. *Inho-ga* *mul-ul* *eol-ly-eot-da*
Inho-NOM water-ACC freeze-CAUS-PAST-IND
'Inho froze the water.'

[1] All cited examples are presented in Revised Romanization. Examples from Yeon (1991), Park (1994), and Kim (2011) were re-transcribed for consistency, but no other changes were made to morphological segmentation or glosses.

In contrast, *ge* (also romanized as *key*) appears in a periphrastic construction with the auxillary verb *ha-ta* 'do.' It is apparently productive in its application, appearing with verbs that disallow *-hi* (4) and many unergative verbs (5). Both constructions are subject to semantic constraints (Park 1994, Kim 2011), but *ge* is not lexically restricted like *-hi*. This makes for a dichotomy between *-hi* and *ge*: *ge* is 'productive' in the sense that it can be applied broadly, while *-hi* is 'non-productive' both in the sense that its distribution is lexically restricted such that the verbs that it can co-occur with must be memorized and in the sense that its form is unpredictable because its allomorphs are not entirely determined by the form of the adjoining root. There is a loose parallel to be made with the English lexical causatives and *make* causatives. Only some English verbs have lexical causatives (*die∼kill*, *eat∼feed*, *fall∼fell*, *rise∼raise*), and their forms are not predictable, while *make* causatives can be productively formed. English lexical causatives have lost productivity over time (alternations such as *fall∼fell* were once more predictable). The same holds for *-hi*: its attested distribution in Middle Korean suggests that it was once more broadly applicable and could even be applied to borrowed vocabulary which is ungrammatical in the modern language. It has become more lexically restricted over time (Park 1994, Shibatani & Pardeshi 2002).

(4) *ge* causative without corresponding *-hi* causative (Park 1994 (65a, c))

a. * *Inho-ga Mina-lul L.A.-e sal-ly-eot-da*
Inho-NOM Mina-ACC L.A.-in live-CAUS-PAST-IND
(Intended: 'Inho made Mina live in L.A.')

b. *Inho-ga Mina-lul L.A.-e sal-ge hae-t-da*
Inho-NOM Mina-ACC L.A.-in live-COMP CAUSE-PAST-IND
'Inho made Mina live in L.A.'

(5) *ge* causative with unergative verb (Shibatani & Pardeshi 2002 (25a-c))

a. *ae-ga chaek-ul ilk-eotda*
child-NOM book-ACC read-PASS.IND
'The child read the book.'

b. *eomma-ga ae-ga/ege/lul chaek-ul ilk-ge*
mother-NOM child-NOM/DAT/ACC book-ACC read-COMP
hae-tda
DO-PAST.INDIC
'Mother made the child read the book.'

The Korean passive alternation is strikingly parallel to the causative both in morphological form and syntactic structure (Kim 2011). It has been argued that the causative *-hi* and passive *-hi* actually constitute polysemous meanings of a single transitivity-altering morpheme (Park 1994). While they are certainly etymologically related and are highly homophonous, it is not entirely clear that they should be analyzed as such. First, there are a few high-frequency verbs for which the causative and passive *-hi* have different allomorphs (*mek-i* 'cause to eat' vs. *mek-hi* 'be eaten'; Yeon 1991). And second, some Korean dialects with lexical tone distinguish the two in that way (Lee 1970; translated in Park (1994; (23))). Both of these

would suggest two synchronically distinct morphemes whose similarities are due to their diachronic relatedness. That said, our acquisition analysis does rest upon a two-morpheme analysis.

3 Learning argument structure generalizations

The acquisition of the causative alternation is a problem of generalization tuning. It is up to each individual child to determine whether such a pattern exists in their native language, whether that pattern is productive, and how extensively that pattern should be applied. Children cannot be expected to arrive at the correct generalizations immediately. One potential outcome of a temporarily over-generalized pattern is over-generalization errors in production. These are readily attested in English, indicating that there exist learners for whom the causative alternation is more productive than it is for adults (Bowerman & Croft 2008, Irani 2019) as in (6).

(6) Over-application of causative alternation in Ross (MacWhinney 1991)

a. Ross: 'I'm want them to disappear them again.'
Intended: 'I want them to make them disappear again.'

b. Ross: 'Maybe I can fall it down the stairs.'
Intended: 'Maybe I can drop it / make it fall down the stairs.'

Recently, the *Sufficiency Principle* (SP; Yang 2016; §6.1.3) has emerged as an input-driven framework for argument structure learning. One advantage of the SP is that it is mechanistic. It makes concrete empirical predictions about the productivity of children's argument structure generalizations given their input. Recent work, Irani (2019), applies the SP to corpora of English learners' input from CHILDES (MacWhinney 2014) to account for the English developmental trajectory: a child who has received limited input is able to hypothesize a grammar in which the causative alternation is productive (hence over-productions), but after the child has received more input, it is no longer productive, so over-production errors cease.

Much less is known about the acquisition of the Korean causatives. First, Choi (1999) performs one of the few formal studies about the development trajectory within the framework of Pinker (1989) and finds a few patterns. Attested "errors" in child productions tend to be under-productions of *-hi* , that is, the use of non-causative forms when an adult would have employed a *-hi* causative. Choi identifies *ge* as productive because of its wide applicability, but also notes that *ge* is acquired later than *-hi* . These are patterns that a model of acquisition should account for.

In addition to this directed study, Park (1994) provides an informal account of the acquisition trajectory of his daughter Ye-hwan. He identifies under-use of *-hi* with examples beginning around 2;4 and lasting until around 3;3 with adult-like use of *-hi* beginning around 2;8 and gradually increasing in regularity. Some of Ye-hwan's anomalous productions (Park 1994; (85a,f)) are provided with context in (7).

(7) Context: 'She sat on a swivel chair and asked me to turn the chair around'

a. Ye-hwan: *bingbing dol-a.jwo*
round turn.around-(CAUS)-BEN.REQ
'Turn around repeatedly'

b. Adult: *bingbing dol-l-yeo.jwo*
round turn.around-(CAUS)-BEN.REQ
'Turn me around repeatedly'

Context: 'She asked me to show a picture to her'

a. Ye-hwan: *bo-a.jwo*
see-BEN.REQ
'Look at it'

b. Adult: *bo-yeo.jwo*
see-CAUS.BEN.REQ
'Show it to me'

In the last several years, corpora of Korean child-directed (CDS) and child-produced speech (CPS) have become available as part of the CHILDES project (MacWhinney 2014) which allow for new empirical investigation of both the input which Korean learners receive and the development of child productions. We describe the input and productions of one child, Yun, within the Ryu corpus (Ryu 2011). Yun's age ranges between 2;3 and 3;9 during the period of recording, and he was chosen because his section of Ryu is very similar in size to Ross's section of the MacWhinney Corpus (MacWhinney 1991), which allows for a degree of comparison between English and Korean (Table 1). Each contains over 80,000 child-directed utterances and over 35,000 child-produced utterances.

Korean Yun	Count	English Ross	Count
Total child-directed utterances	81,577	Total child-directed	82,466
Total child-produced utterances	38,356	Total child-produced	35,912

Table 1: Size of Yun and Ross CDS and CPS in number of utterances

We began by cataloguing all non-adult-like causative uses in Yun's speech in order to compare his productions to Choi's and Park's findings. The results in Table 2 are largely consistent with previous findings, namely, most anomalous productions are the use of non-causatives in place of *-hi* causatives. However, we find a single instance over-application of *-hi* (8) and none for *ge*, and Yun's overall under-application rate is lower than what was described for Ye-hwan who was of similar age. A comparison with Ross reveals an interesting similarity. Errors in Ross were exclusively over-applications of the causative alternation as in (6) in which unaccusative verbs were used transitively. This is reminiscent of under-application of the *-hi* causative in which non-transitive verbs are unexpectedly used in transitive sentences.

However, the mechanism behind the anomalous forms may be different: In English, this may be due to learners generalizing the alternation of unaccusative verbs (e.g., *melt*, *open*) to verbs which should have lexical causatives or *make* causatives (e.g., *disappear*∼*make disappear*, *fall*∼*drop*). This is the over-generalization of a

grammatical pattern. In Korean, however, this may be the under-generalization of the *-hi* causative where the learner has yet to learn which verbs support the construction.

Korean Error Type	Count	English Error Type	Count
Over-application of *-hi* -causative	1	Over-application of alternation	10
Under-application of *-hi* -causative	6	Under-application of alternation	0
Over-application of *ge* -causative	0	-	-
Under-application of *ge* -causative	0	-	-

Table 2: Anomalous causative productions in Yun and Ross

(8) Over-application of *-hi*. Context: Yun is describing a spinning top

a. Yun: *ireoge ige dol-li-neun geo-ya*
how this turn-CAUS-PRES EMPH
'This is how it's turning something'

b. Adult: *ireoge ige do-neun geo-ya*
how this turn-PRES EMPH
'This is how it's turning'

4 Accounting for acquisition

We model child language acquisition as described by the *Sufficiency Principle* (SP; Yang 2016; §6.1.3), which is a corollary to the *Tolerance Principle* (Yang 2016) that has enjoyed recent success as a model of productivity learning, addressing a wide range of problems in syntax. It distinguishes itself as a mechanistic online model by which learners decide whether or not some hypothesized pattern is productive in the grammar. The SP provides a binary outcome: either there is "sufficient" evidence of some pattern in the input to justify productivity, or it is non-productive, meaning that items that appear to obey the pattern are memorized individually. A pattern is productive if enough of the types which a learner knows that could obey the hypothesized pattern actually do. This is determined by a *sufficiency threshold* which is derived given an Elsewhere Condition on the representation of productive patterns (Kiparsky 1973, Aronoff 1976), frequency-rank based lexical access (Murray & Forster 2004), and a generally Zipfian input distribution (Yang 2016; pp. 48-51). A formal definition is given in Equation (1):

Sufficiency Principle: Generalization R applying over N *types* with M attested is *sufficient* if

$$(1) \quad N - M < \theta;\ \theta := \frac{N}{\ln N}$$

More concretely, a Korean learner needs to work out whether the *-hi* and *ge* causatives are productive. If the learner has encountered N potentially causativizable verb types and M of those have been attested in the input with *-hi* , then the learner should postulate that *-hi* is productive if the number of verb types that have

not (yet) been attested $N - M$ is less than the sufficiency threshold $\theta = N/\ln N$. If the construction is productive, then the learner is able to extend it to novel verbs, but if the construction is not productive, then the learner can only use it with the verbs with which it has been attested in his or her input. The same calculation is performed for *ge*.

We calculated sufficiency thresholds on estimations of both child and adult lexicons in order to model the developmental trajectory of causative productivity. Running the calculations over adult grammaticality judgments also serves as a sanity check, since we should expect to find non-productivity in *-hi* and productivity in *ge* in line with existing literature. The lexicons were taken from the CDS within the Yun corpus, which represents the input available to a Korean learner. We categorized the verbs into three groups: unaccusatives, non-unaccusatives (including anything else), and statives. Stative verbs are common in Korean, but rare in English, and they undergo a causative alternation similar to unaccusative verbs (9). Because the stative causative alternation is similar to the unaccusative, it is unclear whether to treat them as a separate class or include them with the unaccusatives. We considered both classifications in our calculations and find that both make the same predictions about productivity. Under the three-way classification, there are $N = 25$ unaccusative verbs, 129 non-unaccusatives, and 74 statives in the Yun CDS lexicon.

(9) Stative causative alternation

a. *dol-yi* *tteugeop-da*
rock-NOM hot-IND
'The rock is hot'

b. *Cheolsu-ga* *dol-eul* *tteugeop-ge hae-t-da*
Chulsoo-NOM rock-ACC hot-CAUS DO-PAST-INDIC
'Chulsoo made the rock hot'

Next, adults' M values for each class were taken from mature native speakers' acceptability judgments. If a given verb from some class can form its causative with either of the constructions, it counted towards the M for that construction. The sufficiency thresholds for each class and both causative constructions were calculated. Table 3 lays out the calculations and results: in each case, *-hi* is non-productive as expected, though it comes much closer for unaccusative verbs ($N - M - \theta = 1.4$) than for the other categories. *Ge*, on the other hand, is productive for all classes since the large majority of verbs support the construction.

Adult Judgments	N	θ	M *-hi*	Productive?	M *ge*	Productive?
unaccusatives	25	7.6	16	$N - M_{hi} = 9$. **no**	25	**yes**
non-unaccusatives	129	26.5	11	$N - M_{hi} = 118$. **no**	128	**yes**
statives	74	17.2	3	$N - M_{hi} = 71$. **no**	66	**yes**

Table 3: Sufficiency Principle calculations over adult judgments

For our child learner, both N and M are both estimated from CDS. The learner is taken to be familiar with a given verb if it appears in the corpus, and is taken to have learned its causative formation only if it is attested. Since a young child's input is limited, the child's M values in Table 4 are much smaller than an adult's. When

the Sufficiency Principle calculations are performed, we find that neither *-hi* nor *ge* is productive for a learner who has received only the input present in Yun's CDS. It should be noted that this estimated learner is not the same as the real Yun since three-year-old Yun certainly received additional input not present in the corpus.

In Yun CDS	N	θ	M *-hi*	Productive?	M *ge*	Productive?
unaccusatives	25	7.6	12	**no**	4	**no**
non-unaccusatives	129	26.5	12	**no**	3	**no**
statives	74	17.2	1	**no**	6	**no**

Table 4: Sufficiency Principle calculations performed over Yun CDS.

This predicts that a child who makes anomalous productions should under-apply either construction rather than over-apply it, since neither construction is productive. This is for the most part consistent with Park (1994), Choi (1999), and our estimations of Yun's CPS. Additionally, there are significantly fewer instances of the *ge* causative attested in Yun than *-hi*. Since the constructions have to be learned verb-by-verb at this stage, this means that such a learner would have access to much fewer *ge* causatives than *-hi* , which may account for Choi's observation that *ge* is acquired later. We expect that *ge* will be rendered productive eventually as the child matures and is exposed to more *ge* causatives over time. As such, the Sufficiency Principle accounts for the attested patterns of Korean language development.

5 Discussion

We perform an analysis of Korean child-directed and child-produced speech in order to better understand the acquisition of the two Korean causative constructions *-hi* and *ge*. We make use of recently available corpora of child-directed and child-produced speech in conjunction with the Sufficiency Principle, a model of productivity learning, to investigate Korean learners' developmental trajectories. We add to what is known about learners' anomalous productions in a corpus study of learner Korean learner Yun's speech in the Ryu corpus. Next, we show that the SP makes correct predictions about the constructions' productivity in adult grammars. Following that, our application of the SP to the lexicon available in Yun predicts the non-productivity of both constructions for a young learner.

Our results account for what is known about the developmental trajectory of Korean causatives. First, *ge* is eventually productive. The construction is clearly productive for adults, so after awhile, a child should hear *ge* with enough unique verbs for it to surpass the sufficiency threshold. However, we find that *ge* is poorly attested in our input sample. Since it is unproductive for early learners, its low attestation means that they have fewer opportunities to learn to use it, which may give the appearance of it being acquired later than *-hi*. Second, *-hi* is unproductive for early learners, thus learners' anomalous productions tend towards the constructions' under-application. Learners who have yet to encounter the construction with a given verb do not know to use it.

The Sufficiency Principle has also been applied to English data to account for learners' over-application of the causative alternation. Both English and Korean learners face Baker's Paradox when acquiring their respective constructions: they need to figure out if and when to generalize a pattern of the grammar on very sparse

evidence. As such, the same learning algorithm applied over input from two different languages accounts for both language's learning trajectories. The acquisition of the Korean and English causative constructions, and Baker's Paradox more generally, exemplifies the dual roles that innate linguistic abilities and language-specific eccentricities play to achieve language acquisition.

References

Aronoff, M. 1976. Word formation in generative grammar. *Linguistic Inquiry Monographs Cambridge, Mass.* 1–134.

Arunachalam, S. 2015. Argument structure: Relationships between theory and acquisition. In *Cognitive science perspectives on verb representation and processing*, 259–280. Springer.

Baker, C. L. 1979. Syntactic theory and the projection problem. *Linguistic Inquiry* 10.533–581.

Bowerman, M., & W. Croft. 2008. The acquisition of the english causative alternation. In *Crosslinguistic perspectives on argument structure: Implications for learnability*, ed. by M. Bowerman & P. Brown, 279–307. Erlbaum.

Choi, J.-H. 1999. Sayektongsa suptuk-ey kwanhan yenkwu ("on the acquisition of causative verbs"). *Journal of Language Science* 6.149–162.

Irani, A. 2019. *Learning from Positive Evidence: The Acquisition of Verb Argument Structure*. University of Pennsylvania dissertation.

Kim, K. 2011. High applicatives in korean causatives and passives. *Lingua* 121.487–510.

Kiparsky, P. 1973. 'Elsewhere' in phonology. In *A festschrift for Morris Halle*, ed. by S. R. Anderson & P. Kiparsky, 93–106. New York: Holt, Rinehart & Winston.

Landau, B., & L. R. Gleitman. 1985. *Language and experience*. Cambridge, MA: Harvard University Press.

Lee, S.-O. 1970. Kwukeuy satong phitong kwumun yenkyu ("a study of causative and passive constructions in Korean"). *Kwukeyenkwu* 26.

MacWhinney, B. 1991. The childes language project: Tools for analyzing talk. *Hillsdale, NJ: Lawrence Erlbaum* 40.62–74.

MacWhinney, B. 2014. *The CHILDES project: Tools for analyzing talk, Volume II: The database*. Psychology Press.

Murray, W. S., & K. I. Forster. 2004. Serial mechanisms in lexical access: the rank hypothesis. *Psychological Review* 111.721–756.

Park, J.-W. 1994. *Morphological causatives in Korean: problems in grammatical polysemy and constructional relations*. University of California, Berkeley dissertation.

Pinker, S. 1984. *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.

Pinker, S. 1989. *Learnability and cognition: The acquisition of argument structure*. MIT press.

Ryu, J.-Y. 2011. The l1 acquisition of the imperfective aspect markers in korean: A comparison with japanese. In *Proceedings of the 25th Pacific Asia Conference on Language, Information and Computation*, 186–195.

Shibatani, M., & P. Pardeshi. 2002. The causative continuum. *Typological studies in language* 48.85–126.

Yang, C. 2016. *The price of linguistic productivity: How children learn to break the rules of language*. MIT Press.

Yeon, J.-H. 1991. The korean causative-passive correlation revisited.