

Acquiring the Korean Causatives.

Sun Jae Lee and Jordan Kodner

Korean has two causative constructions, one morphological, and one periphrastic, that create transitives from unaccusatives. Native learners must determine which constructions each verb supports. We apply the Tolerance Principle (TP) [8], a mechanistic model of productivity learning, to Korean data and find that it correctly predicts the non-productivity of the former and productivity of the latter, behavior consistent with observed child errors. Korean learners' under-production of causative forms can be contrasted with attested English over-production. The same TP model applied to these languages accounts for the different learner behaviors. It also provides an explanation for how evidence in the input compels children to arrive at the adult distribution of forms.

Background: Korean has a productive periphrastic causative construction (*key*-causative) as well as a morphological causative that applies to a fixed set of verbs (*-hi*-causative) [2, 7, 9] for example, *anj-da* 'sit (intrans.)' ~ *anj-hi-da* 'make sit'. The (*-hi*-causative) has some allomorphs (e.g., *sal-da* 'live' ~ *sal-li-da* 'save') whose morphophonological distribution is beyond the scope of this work. The *-hi*-causatives' distribution is only partially semantically predictable. While much has been said about the acquisition of the English causative alternation [1, 4, etc.], little has been said about Korean outside of [2] which identifies three key facts: **1)** Errors are the under-use of *-hi* rather than its over-application, **2)** only *key* is productive, and **3)** the *key*-causative is acquired later than the *-hi*-causative.

The Tolerance Principle is an evaluation metric by which learners determine the productivity of linguistic generalizations. It defines a maximum threshold (the *tolerance threshold*) for the number of exceptions above which a generalization becomes untenable. If the generalization is tenable, it is productive, and the exceptions can be learned lexically. If untenable, then the generalization is not productive. A generalization that could apply to N items (number of types, not token frequency), and with e exceptions to that generalization is said to be tolerable if $e < N / \ln N$. Experimentally, it makes accurate predictions about how children detect productivity from distributional data [6, 8].

It is possible for learners with small vocabularies to temporarily hypothesize productivity for generalizations that are non-productive for adults and vice-versa. Children in such states may occasionally produce over- and under-production errors respectively. As noted in [1, 4, etc.], English learners tend to over-produce the causative alternation (e.g., "*I died him*"), which corresponds to an outsized number of alternators (verbs that can be used in some form both as an unaccusative and as a transitive e.g., *open*, *melt*) in child-directed speech and suggests its productivity given those learners' vocabularies. This contrasts with Korean, where the *-hi*-causative is under-produced and *key* is slightly over-produced (by adult native Korean speaker judgment). The below table contrasts the causative error types in child-produced utterances in the Yun (Korean Ryu [5]) and Ross (Brown) corpora. The corpora are comparable in size, as shown in the same table.

Experiment: We use the TP to calculate the productivity of the Korean causative constructions for adults and learners to relate this to child production errors. For adult speakers, we

Korean CPS Error Type	Count	English CPS Error Type	Count
Overuse of <i>-hi</i> -causative	1	Overuse of alternation	10
Underuse of <i>-hi</i> -causative	6	Underuse of alternation	0
Overuse of <i>key</i> -causative	2	-	-
Underuse of <i>key</i> -causative	0	-	-
Total child-directed utterances	81,577	Total child-directed	82,466
Total child-produced utterances	38,356	Total child-produced	35,912

extract all verbs from the CDS in the Korean CHILDES Ryu (Yun) corpus [5] and classify each verb according to native judgments as an *alternator* (that undergoes the unaccusative alternation in some form) or *non-alternator* (that can still be made causative like *see~make see*) along the lines of [3, 4]'s semantic classifications. We also included stative verbs as a third class since some undergo an alternation (e.g., *be red~make red*). There were two reasons to use CDS here: First, it allows us to compare directly with child learners below. It shows us how children's limited experience compared to adults can drive them to reach (temporarily) different conclusions. Second, CDS contains mostly high frequency common verbs, the kind

that everyone is likely to learn as children when they are sorting out the rules of their grammar.

The *alternator/non-alternator/stative* classification defines the domain of generalization for the TP. We calculate whether either causative construction is productive for each class. N is the size of each class in number of types, and e is the number of members which do not form causatives by the construction being tested (i.e., exceptions to the generalization). We run the calculation for each class and find that the results correspond to our expectations of the adult grammar: *-hi* is predicted to be non-productive for any class (it is lexically specific), and *key* is predicted to be productive for each by the TP in the following table. Collapsing *stative* into either of the other classes does not affect the result.

(Adult judgments)	<i>-hi</i> -Caus	<i>key</i> -Caus	Total (N)	TP Threshold	<i>-hi</i> Productive?	<i>key</i> Productive?
Alternators	16	25	25	$25/\ln 25 = 7.8$	$9 > 7.8$. no	$0 < 7.8$. yes
Non-alternators	11	128	129	26.5	$118 > 26.5$. no	$1 < 26.5$. yes
Statives	3	66	74	17.2	$71 > 17.2$. no	$8 < 17.2$. yes

Next, we identify those verbs actually attested in causative constructions in the child-directed (CDS) part of Yun to estimate the evidence on which a young learner would base their judgments of productivity. When calculating the TP this time, we find that neither construction should be productive for a learner exposed to this as input. The contrast is due to the low attestation of *key*-causative in the CDS. Even though *key* is eventually rendered productive for adults, there is insufficient evidence in this corpus for a learner to reach that conclusion. The following table shows these results.

(in Yun CDS)	<i>-hi</i> -Caus	<i>key</i> -Caus	Total (N)	TP Threshold	<i>-hi</i> Productive?	<i>key</i> Productive?
Alternators	12	4	25	7.8	$13 > 7.8$. no	$21 > 7.8$. no
Non-alternators	12	3	129	26.5	$117 > 26.5$. no	$126 > 26.5$. no
Statives	1	6	74	17.2	$73 > 17.2$. no	$68 > 17.2$. no

Analysis: The Tolerance Principle calculations here account for the observed pattern of over- and under-use in Korean child-produced utterances. **1)** Morphological *-hi* errors are predominantly under-use because it is not productive under young learner or adult lexicons. If it is not productive, then children must learn for each verb whether they can use it. **2)** *key* is productive for adults. We can see that at some point in development, a mature learner has heard *key*-causatives with enough verb types to allow for productivity as predicted by the TP. **3)** *key* is acquired later than *-hi*. We see that under our estimation of a child input, the *key* causative is still unproductive, so it needs to be learned lexically on a word-by-word basis like *-hi* at this point. We also note that there are only about half as many attested *key*-causatives compared to *-hi*, suggesting that fewer *key*-taking lexical items are available to learn. Since both are learned on a word-by-word basis, this provides more opportunity to use the *-hi*-causative than *key*, giving the appearance of later acquisition. We expect that over time, enough examples of *key* are racked up to allow its productivity since adults can use the construction freely. The couple examples of *key* over-production in the child-produced section suggest that Yun himself is beginning to reach productivity at the time of the corpus data collection.

- [1] M. Bowerman and W. Croft. 2008. The acquisition of the English causative alternation. [2] J.-H. Choi. 1999. Sayektongsa suptuk-ey kwanhan yenkwu (on the acquisition of causative verbs). [3] B. Levin. 1993. *English verb classes and alternations*. [4] S. Pinker. 1989. Learning, development, and conceptual change. learnability and cognition: The acquisition of argument structure. [5] J.-Y. Ryu. Ryu corpus: <https://childes.talkbank.org/access/eastasian/korean/ryu.html>. [6] K. Schuler, C. Yang, and E. Newport. 2016. Testing the tolerance principle: Children form productive rules when it is more computationally efficient to do so. [7] M. Shibatani and P. Pardeshi. 2002. The causative continuum. [8] C. Yang. 2016. *The price of linguistic productivity: How children learn to break the rules of language*. [9] J.-H. Yeon. 1991. The Korean causative-passive correlation revisited.