CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Korean-speaking children's constructional knowledge about a transitive event: Corpus analysis and Bayesian modelling

Gyu-Ho SHIN[1,]* (ID) and Seongmin MUN[2]

[1]Department of Asian Studies, Palacký University Olomouc, tř. Svobody 26, 779 00 Olomouc, Czech Republic
[2]Department of English Language and Literature, Chosun University, 309, Pilmun-daero, Dong-gu, Gwangju, 61452, Republic of Korea
*Corresponding author. Gyu-Ho Shin, Department of Asian Studies, tř. Svobody 26, 779 00 Olomouc, Czech Republic. Email: gyuho.shin@upol.cz

Abstract

We investigate Korean-speaking children's knowledge about clause-level constructions involving a transitive event – active transitive and suffixal passive – through corpus analysis and Bayesian modelling. The analysis of Korean caregiver input and children's production in CHILDES revealed that the rates of constructional patterns produced by the children mirrored those uttered by the caregivers to a considerable degree and that the caregivers' use of case-marking was skewed towards single form-function pairings (despite the multiple form-function associations that the markers manifest). Based on these characteristics, we modelled a Bayesian learner by employing construction-based input (without considering lexical information). This simulation revealed the dominance of several constructional patterns, occupying most of the input, and their inhibitory effects on the development of the other patterns. Our findings illuminate how children shape clause-level constructional knowledge in Korean, an understudied language for this topic, as a function of input properties and domain-general learning capacities, appealing to the usage-based constructionist approach.

Keywords: corpus analysis; Bayesian simulation; constructional knowledge; Korean

Introduction

A usage-based constructionist approach assumes that the development of linguistic knowledge occurs via interactions between exposure to linguistic environments and more basic forces from cognitive–psychological factors (e.g., Ambridge, Kidd, Rowland & Theakston, 2015b; Goldberg, 2019; Lieven, 2010; Tomasello, 2003). Linguistic knowledge exists as clusters of form–function pairings (i.e., constructions; Goldberg, 1995), with varying degrees of abstraction. The emergence and growth of these clusters are affected by diverse factors, such as the distributional properties of individual items (e.g., Abbot-Smith & Tomasello, 2006; Tomasello, 2003), the nature of the form–function mapping of each

CrossMark

item (e.g., Cameron-Faulkner, Lieven & Theakston, 2007), the degree of (in)consistency involving the current stimulus against prior experience (e.g., Dittmar, Abbot-Smith, Lieven & Tomasello, 2008), and domain-general learning mechanisms (e.g., Langacker, 2017; Stefanowitsch, 2011; Theakston, 2004). As learning occurs, some of the clusters are strengthened enough to reliably defeat other competitors (Bates & MacWhinney, 1982, 1989; MacWhinney, 1987; Goldberg, 2019). Existing evidence, most of which comes from Indo-European languages (e.g., Aguado-Orea & Pine, 2015; Behrens, 2006; Cameron-Faulkner, Lieven & Tomasello, 2003; Ibbotson & Tomasello, 2009), supports the core assumption of this approach that ascribes the development of linguistic knowledge to the interplay between input properties and domain-general learning capacities.

One issue within this approach is how to better capture developmental trajectories of children's linguistic knowledge based on the exposure that they receive. Recently, researchers have used computational modelling to address this issue. As a proxy for the conceptual space in human cognition, a simulation provides a reliable model for how learning occurs (e.g., Ambridge & Blything, 2016; Ambridge et al., 2020; Lupyan & Christiansen, 2002; Matusevych, Alishahi & Backus, 2016). Specifically, emerging research has shown the effectiveness of Bayesian inference for this kind of task (e.g., Alishahi & Stevenson, 2008; Bannard, Lieven & Tomasello, 2009; Barak, Goldberg & Stevenson, 2016; Barak & Goldberg, 2017; Nguyen & Pearl, 2019; Perfors, Tenenbaum, Griffiths & Xu, 2011a; Xu & Tenenbaum, 2007), assuming that human learning involves updating one's beliefs based on previous experience. However, there is one caveat to this practice: previous research has been skewed heavily towards English, so it is uncertain to what degree the implications of these simulation studies are generalisable across languages to support the core assumption of the usage-based constructionist approach.

Against this background, the present study explores how Korean-speaking children develop their knowledge about representative argument structure constructions expressing a transitive event – an active transitive and a suffixal passive – as a function of input properties and statistical learning. This proceeds in two ways. One is the analysis of caregiver input and child production in CHILDES (MacWhinney, 2000), the largest open-access child corpora in Korean. The other is a Bayesian simulation that employs information about the frequency of the two construction types in the corpus. Korean, an understudied language in this regard, provides an interesting testbed because of language-specific properties such as agglutination and scrambling/omission of sentential components, which are distinct from characteristics of most languages that have thus far been studied in this regard. Some studies have analysed Korean-speaking children's acquisition of these constructions through behavioural experiments (e.g., Jin, Kim & Song, 2015; Kim, Sung & Yim, 2017; Shin, 2020). However, we are not aware of any study on their developmental trajectories involving clause-level constructions by combining corpus analysis and computational modelling through the window of the usage-based constructionist approach.

Active transitive and suffixal passive in Korean

Korean is an agglutinative, Subject–Object–Verb language[1] with overt case-marking. These structural cues allow scrambling of pre-verbal arguments if that reordering

---

[1]Sometimes, it is possible to place arguments post-verbally: "In colloquial speech, the predicate-final constraint is often relaxed, with some non-predicate elements being uttered after the predicate for

preserves the original intention with no ambiguity. Korean also permits the omission of almost all sentential elements: as long as participants in an event are clearly identified within that context, a case marker or a combination of an argument and a case marker can be omitted without changing the basic propositional meaning.

A canonical active transitive construction (1a) typically occurs with the nominative-marked agent, followed by the accusative-marked theme. The thematic roles of each argument are indicated by designated case markers: a nominative case marker (NOM) *-i/ka* (*-i* after a consonant) and an accusative case marker (ACC) *-(l)ul* (*-ul* after a consonant). The two arguments can be scrambled, comprising the theme–agent ordering (1b). In addition, a case marker (2a) or a noun and a case marker altogether (2b) can be omitted where relevant.

(1a)　Active transitive: canonical
　　　Chelswu-ka　　Yenghuy-lul　　an-ass-ta.
　　　Chelswu-NOM　Yenghuy-ACC　hug-PST-SE[2]
　　　'Chelswu hugged Yenghuy.'

(1b)　Active transitive: scrambled
　　　Yenghuy-lul　　Chelswu-ka　　an-ass-ta.
　　　Yenghuy-ACC　Chelswu-NOM　hug-PST-SE
　　　'Chelswu hugged Yenghuy.'

(2a)　Omission: case marker
　　　Chelswu-ka　　Yenghuy-~~lul~~　　an-ass-ta.
　　　Chelswu-NOM　Yenghuy-~~ACC~~　hug-PST-SE
　　　'Chelswu hugged Yenghuy.'

(2b)　Omission: argument + case marker
　　　~~Chelswu-ka~~　　Yenghuy-lul　　an-ass-ta.
　　　~~Chelswu-NOM~~　Yenghuy-ACC　hug-PST-SE
　　　'(Chelswu) hugged Yenghuy.'

Previous research on Korean-speaking children's acquisition of the active transitive shows an asymmetry by canonicity. To illustrate, the canonical pattern is employed more reliably than its scrambled counterpart (e.g., Jin et al., 2015; Kim et al., 2017; Shin, 2021). Children also tend to map the initial noun onto the agent until the age of four, regardless of its actual thematic role (e.g., Kim, O'Grady & Cho, 1995; No, 2009). This is consistent with the oft-mentioned AGENT-FIRST strategy, which is found in many languages (e.g., Abbot-Smith, Chang, Rowland, Ferguson & Pine, 2017; Huang, Zheng, Meng & Snedeker, 2013; but see Garcia & Kidd, 2020; Shin, 2021).

'after-thought' clarification, amplification of information, or emphasis." (Sohn, 1999, p. 295). This study considers only verb-final sentences hereafter.

[2] Abbreviation: ACC = accusative case marker; DAT = dative marker; NOM = nominative case marker; PSV = passive suffix; PST = past tense marker; SE = sentence ender.

A canonical suffixal passive construction[3] (3a) occurs with the NOM-marked theme, followed by the dative-marked agent indicated by a dative marker (DAT)[4] *-eykey/ hanthey.* The verb carries dedicated passive morphology as one of the four suffixes: *-i, -hi, -li,* and *-ki* (under allomorphic distribution). This pattern can be scrambled, yielding the agent–theme ordering, (3b). The same kind of omission as in (2a–b) also occurs where relevant.

    (3a)   Suffixal passive: canonical
                Yenghuy-ka   Chelswu-hanthey   an-ki-ess-ta.
                Yenghuy-NOM   Chelswu-DAT   hug-PSV-PST-SE
                'Yenghuy was hugged by Chelswu.'

    (3b)   Suffixal passive: scrambled
                Chelswu-hanthey   Yenghuy-ka   an-ki-ess-ta.
                Chelswu-DAT   Yenghuy-NOM   hug-PSV-PST-SE
                'Yenghuy was hugged by Chelswu.'

Contrary to the case of the active transitive, how Korean-speaking children acquire the suffixal passive is inconclusive. Children up to four years of age are not generally adept at the passive in Korean (e.g., Kim et al., 2017; Shin, 2020), which aligns with the attested difficulties with passives cross-linguistically (e.g., Abbot-Smith et al., 2017; Huang et al., 2013). However, their performance diverges after the age of four depending upon task types and verb types. For example, five-and-six-year-old children perform at chance-level comprehension (Kim et al., 2017; Shin, 2020), but their production of the passive could be primed (Kim, 2010). Verb semantics also seems to affect their comprehension so that five-year-olds show above-chance performance in accomplishment verbs but at-chance performance in stative verbs (Lee & Lee, 2008). Hence, these mixed reports make it difficult to gain a clear understanding of children's developmental trajectory involving the passive.

We identify three language-specific aspects that may render the learning process of the two construction types in Korean particularly challenging for children. First, the form–function associations involving case-marking dedicated to these constructions are not straightforward (e.g., Choo & Kwak, 2008). This is particularly true for the two markers, NOM and DAT. For example, whereas NOM primarily indicates a nominal that designates the instigator of an action (i.e., the agent role), as in (1a–b), the same marker indicates the theme in the passive, as in (3a–b). In a similar vein, DAT basically indicates that a nominal is a recipient (in a ditransitive construction), but this marker indicates the agent in the passive as in (3a–b). Therefore, this aspect could affect how the children acquire knowledge about case-marking and clause-level constructions in which the markers engage. Children generally use these markers from the age of two or three, but their understanding of case-marking is not complete until the age of four (e.g., Cho, 1982;

---

[3]There are three types of passives in Korean: suffixal, lexical, and periphrastic (Sohn, 1999; but see Yeon, 2015). All three passive types are rare in the input, but amongst them, lexical and periphrastic passives are extremely rare. We thus focus on the suffixal passive, the representative passive type that children are most likely to encounter. See Shin (2020) for more details on this point via a comprehensive analysis of CHILDES.

[4]For the sake of consistency, we classify DAT as a type of case-marking.

Chung, 1994; Lee, Kim & Song, 2013; No, 2009). However, few attempts have been made to precisely address the impact of the multiple form–function mapping of case-marking on the development of constructional knowledge.

Second, case markers exhibit varying degrees of omission. The optionality of these markers, particularly of NOM and ACC, is observed in colloquial speech (Sohn, 1999); compared to NOM, ACC tends to be occasionally dropped (Chung, 1994). This characteristic seems to affect the acquisition of knowledge about NOM and ACC within the active transitive. Evidence shows that children learn NOM as an indicator of the subject in a sentence as early as 18 to 20 months old (e.g., Cho, 1982; Lee, 2004) and they typically employ a NOM-marked argument as the agent of an event (Kim, 1997; Lee & Cho, 2009; No, 2009). Notably, children acquire NOM earlier and use it more reliably than they use ACC (e.g., Jin et al., 2015; Kim et al., 2017; Lee et al., 2013), which suggests an asymmetry regarding the developmental order of these markers. What remains to be discovered is how this asymmetric nature of case-marking omission influences children's acquisition of the active transitive.

Third, verbal morphology serves as a core element in the passive: only this suffix indicates that a sentence is in the passive voice, signalling that the NOM-marked argument is not the agent but the theme and that the DAT-marked argument is the agent instead. Therefore, the sensitivity to passive morphology is crucial for successful acquisition of the passive in Korean (e.g., Shin, 2020). However, this morphology rarely occurs in input due to the scarcity of the passive in usage. It is also morphologically irregular (e.g., Yeon, 2015) and unproductive because it applies only to a limited set of verbs (e.g., Lee & Lee, 2008; Sohn, 1999). Beyond this, it overlaps with verbal morphology used for a morphological causative construction (e.g., Sohn, 1999). Despite these challenges to the acquisition of the Korean suffixal passive induced by verbal morphology, most previous studies have shown only age factors in acquiring the passive, with the role of verbal morphology unexplored.

With these in mind, we investigate (i) the linguistic environments surrounding Korean-speaking children pertaining to transitive events and (ii) their acquisition of the two construction types as a function of input properties (centring around these constructions) and statistical learning (Bayesian inference). In the next section, we probe into the first inquiry by presenting an analysis of Korean child corpora in CHILDES as an exploratory study.

Analysis of caregiver input and child production

Methods

We analysed all the Korean child corpora available in CHILDES. The dataset consists of 81,593 sentences from nine caregivers and 38,388 sentences from four children whose ages range from 1;3 to 3;10 (Table 1). Of primary interest in this analysis were the active transitive (1a–b) and suffixal passive (3a–b), with or without the omission of such obligatory components as arguments and case-marking, with (non-)canonical word order. We also examined the use of individual markers – NOM, ACC, and DAT – dedicated to the two construction types.

CLAN, a default programme of CHILDES for data analysis and editing, is not supported for Korean, so the analysis was conducted through *Python* programming in a semi-automatic way. As the raw data were not suitable for an automatic pattern-finding

Table 1. Information about corpora.

Name of corpus	Caregiver	Child / age range	Time of collection (year)	Quantity (sentence #)	
				Caregiver	Child
Jiwon	M & F	Jiwon / 2;0–2;3	1992	10,602	6,443
Ryu	GM, GF & M	Jong / 1;3–3;5	2009–2011	28,657	13,698
	GM, M & F	Joo / 1;9–3;10	2010–2011	27,071	11,730
	M	Yun / 2;3–3;9	2009–2010	15,263	6,517

Note. F = father; GM = grandmother; GF: grandfather; M = mother.

process, they were applied first to a pre-processing stage: typos and spacing errors were corrected; part-of-speech tagging information was attached automatically and revised manually; lines whose length was less than five strings (i.e., characters) or those consisting only of onomatopoeia and mimetic words were excluded (see Shin, 2020 for the details about the pre-processing). Any non-verb-final instance (e.g., *Yengswu*-NOM read-SE book-ACC; eat-SE rice-ACC) was also excluded from the data. These treatments resulted in 69,498 sentences (285,350 eojeols[5]) for the caregiver input and 1,985 sentences (25,047 eojeols) for the children's production.

Next, the pre-processed data were inputted to an automatic search process whereby instances of the two construction types and the three markers involving these constructions were extracted. To illustrate, the canonical active transitive with no omission was identified through the following steps. First, we isolated instances with a verb and more than one noun. Second, of these instances, we extracted cases with both NOM and ACC. Lastly, from these cases, we outputted sentences where NOM preceded ACC as a text file. Beyond these steps, every list of sentences for each extraction was checked manually to ensure its accuracy.

In addition to the frequency information about each pattern and case-marking, we calculated ΔP, a unidirectional statistics for association strength that estimates the degree to which a cue co-occurs with an outcome (e.g., Allan, 1980; Desagulier, 2016). A ΔP score, which ranges from –1 to 1, is computed based on a contingency table (Table 2), following the mathematical formula (4) where the probability of the outcome is conditioned upon that of the cue. For the interpretation of ΔP scores, the closer $\Delta P_{(outcome|cue)}$ is to 1, the more likely the cue co-occurs with the outcome; the closer $\Delta P_{(outcome|cue)}$ is to –1, the more unlikely the cue co-occurs with the outcome. We applied this technique to the individual markers used in the two construction types to discover how these markers invite the

Table 2. Association strength: ΔP (¬ stands for 'not').

	Outcome	¬ Outcome
Cue	a	b
¬ Cue	c	d

---

[5]An *eojeol* is defined as a unit with white space on both sides that serves as the minimal unit of sentential components. It corresponds roughly to what we call a (tokenised) word in English.

corresponding thematic roles and vice versa in the target constructional patterns (cf. Ramscar, Yarlett, Dye, Denny & Thorpe, 2010).

$$\Delta P = p(\text{Outcome}|\text{Cue}) - p(\text{Outcome}|\neg\text{Cue}) = a/(a+b) - c/(c+d) \qquad (4)$$

Results: caregiver input

Construction

Table 3 presents frequency information about all the possible constructional patterns for a transitive event in the caregiver input. There were five major trends in the caregiver input.[6] First, the number of first-noun-as-agent patterns (3,049 instances) did not exceed that of first-noun-as-theme patterns (3,579 instances). Second, short, simple utterances (e.g., one-argument patterns; 4,561 instances) occurred more frequently than two-argu-ment patterns (2,107 instances). Third, the passive patterns were rare in the input (443 instances) compared to the active ones (6,225 instances), but the number of passive patterns with only one case-marked argument was relatively large (420 instances). Fourth, within the active transitive, once two arguments were attested overtly, most of the utterances followed the canonical word order (i.e., agent-before-theme; 2,047 out of 2,104 instances). Fifth, within the active patterns, the omission rate of NOM (0.01) was considerably lower than that of ACC (0.23).

Case-marking

NOM involves two functions for the two construction types, indicating either the agent (for the active transitive) or the theme (for the suffixal passive). Table 4 presents frequency information about NOM based on the thematic role associated with it and whether/where the case-marked argument appeared in the patterns extracted from the caregiver input. NOM was used more as an indication of the agent than an indication of the theme. The ΔP scores substantiated the strong bi-directional association between NOM and the agent role in the context of a transitive event. NOM and the agent were extremely reliable cues for each other ($\Delta P_{(\text{AGENT}|\text{NOM})} = 0.853$; $\Delta P_{(\text{NOM}|\text{AGENT})} = 0.856$). In contrast, NOM was highly unlikely to introduce the theme ($\Delta P_{(\text{THEME}|\text{NOM})} = -0.868$) and vice versa ($\Delta P_{(\text{NOM}|\text{THEME})} = -0.905$). This indicates strong cue validity (cf. Bates & MacWhinney, 1989) between NOM and the agent within the transitive-event-related constructional patterns in the caregiver input.

Within the active transitive, ACC typically indicates the theme. Table 5 presents frequency information about ACC based on whether/where the case-marked argument appeared in the patterns extracted from the caregiver input. Considering the overall frequency of the patterns in Table 3, the number of the ACC-related patterns was relatively large. In particular, the one-argument pattern with only ACC present (1,938 instances) occurred as frequently as the other two patterns (51 + 1,776 = 1,827 instances), which yielded no statistical significance: $\chi^2(1) = 2.924$, $p = .087$. The ΔP scores showed that the association between ACC and the theme role within a transitive event was moderately reliable. That is, ACC was a dependable cue for the theme ($\Delta P_{(\text{THEME}|\text{ACC})} = 0.350$) and vice versa ($\Delta P_{(\text{ACC}|\text{THEME})} = 0.670$) but not extremely

---

[6]We additionally checked the individual patterns of input by caregivers but there was no meaningful tendency in those patterns.

Table 3. Constructional patterns (with or without scrambling and omission of sentential components) for a transitive event in the caregiver input (adapted from Shin, 2020).

Construction		Example	Frequency #	Frequency %
Canonical active transitive	No omission	Chelswu-NOM Yenghuy-ACC hug	1,757	25.46
	no ACC	Chelswu-NOM Yenghuy-~~ACC~~ hug	268	3.88
	no NOM	Chelswu-~~NOM~~ Yenghuy-ACC hug	19	0.28
Scrambled active transitive	No omission	Yenghuy-ACC Chelswu-NOM hug	51	0.74
	no NOM	Yenghuy-ACC Chelswu-~~NOM~~ hug	0	0.00
	no ACC	Yenghuy-~~ACC~~ Chelswu-NOM hug	6	0.09
Active Transitive with omission	agent–theme, no CM	Chelswu-~~NOM~~ Yenghuy-~~ACC~~ hug	3	0.04
	theme–agent, no CM	Yenghuy-~~ACC~~ Chelswu-~~NOM~~ hug	0	0.00
	undetermined, no CM	Chelswu-~~NOM~~ Yenghuy-~~ACC~~ hug	0	0.00
	agent–NOM only[1]	Chelswu-NOM hug	935	13.55
	theme–ACC only[1]	Yenghuy-ACC hug	1,938	28.08
	agent only, no CM[1]	Chelswu-~~NOM~~ hug	53	0.77
	theme only, no CM[1]	Yenghuy-~~ACC~~ hug	1,155	16.73
	undetermined, no CM[1]	Chelswu-~~NOM~~ hug	40	0.58
Canonical suffixal passive	No omission	Yenghuy-NOM Chelswu-DAT hug-psv	2	0.03
	no DAT	Yenghuy-NOM Chelswu-~~DAT~~ hug-psv	0	0.00
	no NOM	Yenghuy-~~NOM~~ Chelswu-DAT hug-psv	0	0.00
Scrambled suffixal passive	No omission	Chelswu-DAT Yenghuy-NOM hug-psv	1	0.01
	no NOM	Chelswu-DAT Yenghuy-~~NOM~~ hug-psv	0	0.00
	no DAT	Chelswu-~~DAT~~ Yenghuy-NOM hug-psv	0	0.00
Suffixal passive with omission	theme–agent, no CM	Yenghuy-~~NOM~~ Chelswu-~~DAT~~ hug-psv	0	0.00
	agent–theme, no CM	Chelswu-~~DAT~~ Yenghuy-~~NOM~~ hug-psv	0	0.00
	undetermined, no CM	Yenghuy-~~NOM~~ Chelswu-~~ACC~~ hug-psv	0	0.00
	theme–NOM only[1]	Yenghuy-NOM hug-psv	407	5.90
	agent–DAT only[1]	Chelswu-DAT hug-psv	13	0.19
	theme only, no CM[1]	Yenghuy-~~NOM~~ hug-psv	20	0.29
	agent only, no CM[1]	Chelswu-~~DAT~~ hug-psv	0	0.00
	undetermined, no CM[1]	Chelswu-~~NOM~~ hug-psv	0	0.00
Ditransitive	recipient–DAT only[2]	Chelswu-DAT give	234	3.39
	SUM		6,902	100.00

Note. CM = case-marking. 1) does not involve canonicity as it is undeterminable with only one overt argument. Although 2) does not relate to a transitive event *per se* and does not count as a relevant pattern, we considered it here because DAT is often used to indicate a recipient in the active and thus a potential competitor of the agent–DAT pairing in the passive.

Table 4. Frequency of NOM in caregiver input.

Thematic role	Appeared?	Where?	Type	Frequency (#)
Agent	Yes	Initially	One-argument	935
			Two-argument, canonical	2,025
		Non-initially	Two-argument, scrambled	57
	No	Initially	One-argument	53
			Two-argument, canonical	22
		Non-initially	Two-argument, scrambled	0
Theme	Yes	Initially	One-argument	407
			Two-argument, canonical	2
		Non-initially	Two-argument, scrambled	1
	No	Initially	One-argument	20
			Two-argument, canonical	0
		Non-initially	Two-argument, scrambled	0

Table 5. Frequency of ACC in caregiver input.

Thematic role	Appeared?	Where to appear?	Type	Frequency (#)
Theme	Yes	Initially	One-argument	1,938
			Two-argument, scrambled	51
		Non-initially	Two-argument, canonical	1,776
	No	Initially	One-argument	1,155
			Two-argument, scrambled	6
		Non-initially	Two-argument, canonical	271

Note. Since the focus of analysis was patterns involving a transitive event, we excluded any ditransitive pattern.

strong as occurred in the case of NOM and the agent. This was caused by the high omission rate for ACC compared to that of NOM, by increasing the impact of the ¬ cue on calculation of ΔP. These results indicate that, despite the consistent mapping between form and function within the active transitive, the theme–ACC pairing manifests weaker cue validity (cf. Bates & MacWhinney, 1989) than the agent–NOM pairing, particularly when ACC invites the theme.

For DAT, there were only 16 instances where this marker indicated an agent in the passive. The ΔP scores further revealed that DAT and the agent were unlikely to be associated with each other ($\Delta P_{(\text{AGENT}|\text{DAT})} = -0.507$; $\Delta P_{(\text{DAT}|\text{AGENT})} = -0.098$). Although the active patterns involving DAT are mostly ditransitives (and therefore do not count as relevant patterns expressing transitive events), we considered them here because DAT, often used as an indicator of a recipient in the active, serves as a potential competitor of the agent–DAT pairing in the passive. Together with the low frequency of the agent–DAT

pairing, this attribute considerably aggravates the cue validity (cf. Bates & MacWhinney, 1989) of DAT for the agent role (consequently, this pairing gives way to the agent–NOM pairing).

Result: child production

Table 6 presents frequency information about all the constructional patterns for a transitive event in the children's production.[7] When expressing a transitive event, the children utilised only a few patterns: the canonical active transitive with no omission (37 instances), the active transitive with the no-ACC theme argument (30 instances), and the one-argument active patterns with either the theme–ACC pairing (25 instances) or the agent–NOM pairing (21 instances). Of the two-argument patterns with case-marking omitted, the children used only the no-ACC pattern (14 instances). There were nine instances of the one-argument passive pattern with the theme–NOM pairing: of these instances, four included the verb po-i- 'see-PSV' and two included the verb yel-li- 'open-PSV' (beyond these, we could not find such skewness in the rest of the patterns that the children uttered).

Discussion

Although the amount of data for the two construction types was small (9.93% and 5.34% for the entire data of the caregiver input and children's production, respectively), thus requiring cautious interpretation, this analysis yielded two major findings.

First, we found an asymmetry in the frequency of the two construction types for expressing transitive events. In the caregiver input, the active transitive occupied most of the input composition, but there were generally more theme-first patterns than agent-first patterns. There were more instances of one-argument patterns than those of two-argument patterns, reflecting the general characteristic of caregiver input (e.g., Cameron-Faulkner et al., 2003). Regarding the two-argument patterns, the canonical pattern occurred more frequently than the scrambled pattern.

The asymmetries involving the constructional patterns in the caregiver input were induced by such factors as thematic role ordering, the number of arguments, voice type, and the omission of sentential elements. These factors appear to manipulate the degree to which each pattern was available or reliable when the children acquired constructional knowledge from the caregiver input (Bates & MacWhinney, 1982, 1989; MacWhinney, 1987). Indeed, we found that the characteristics of the children's production generally mirrored those of the caregiver input. For instance, the children employed the canonical active transitive as the core construction type for expressing a transitive event, which was also the dominant pattern in the caregiver input. The children's use of patterns that included omissions also resembled the same tendency found in the caregiver input. This aligns with previous literature highlighting the direct connection between caregiver input and children's development of linguistic knowledge (e.g., Ambridge et al., 2015b; Behrens, 2006; Cameron-Faulkner et al., 2003; Stoll, Abbot-Smith & Lieven, 2009).

---

[7]We also checked the individual patterns of production by children. There was no meaningful tendency except that one of the four children did not produce the suffixal passive with the theme–NOM only pattern. However, considering the extremely small number of overall occurrences (nine instances), further investigation does not seem warranted.

Table 6. Constructional patterns (with or without scrambling and omission of sentential components) for a transitive event in child production.

	Construction	Example	Frequency #	%
Canonical active transitive	No omission	Chelswu-NOM Yenghuy-ACC hug	37	25.87
	no ACC	Chelswu-NOM Yenghuy-~~ACC~~ hug	14	9.79
	no NOM	Chelswu-~~NOM~~ Yenghuy-ACC hug	0	0.00
Scrambled active transitive	No omission	Yenghuy-ACC Chelswu-NOM hug	0	0.00
	no NOM	Yenghuy-ACC Chelswu-~~NOM~~ hug	0	0.00
	no ACC	Yenghuy-~~ACC~~ Chelswu-NOM hug	0	0.00
Active transitive with omission	agent–theme, no CM	Chelswu-~~NOM~~ Yenghuy-~~ACC~~ hug	0	0.00
	theme–agent, no CM	Yenghuy-~~ACC~~ Chelswu-~~NOM~~ hug	0	0.00
	undetermined, no CM	Chelswu-~~NOM~~ Yenghuy-~~ACC~~ hug	0	0.00
	agent–NOM only[1]	Chelswu-NOM hug	21	14.69
	theme–ACC only[1]	Yenghuy-ACC hug	25	17.48
	agent only, no CM[1]	Chelswu-~~NOM~~ hug	1	0.70
	theme only, no CM[1]	Yenghuy-~~ACC~~ hug	30	20.98
	undetermined, no CM[1]	Chelswu-~~NOM~~ hug	1	0.70
Canonical suffixal passive	No omission	Yenghuy-NOM Chelswu-DAT hug-psv	0	0.00
	no DAT	Yenghuy-NOM Chelswu-~~DAT~~ hug-psv	0	0.00
	no NOM	Yenghuy-~~NOM~~ Chelswu-DAT hug-psv	0	0.00
Scrambled suffixal passive	No omission	Chelswu-DAT Yenghuy-NOM hug-psv	0	0.00
	no NOM	Chelswu-DAT Yenghuy-~~NOM~~ hug-psv	0	0.00
	no DAT	Chelswu-~~DAT~~ Yenghuy-NOM hug-psv	0	0.00
Suffixal passive with omission	theme–agent, no CM	Yenghuy-~~NOM~~ Chelswu-~~DAT~~ hug-psv	0	0.00
	agent–theme, no CM	Chelswu-~~DAT~~ Yenghuy-~~NOM~~ hug-psv	0	0.00
	undetermined, no CM	Yenghuy-~~NOM~~ Chelswu-~~ACC~~ hug-psv	0	0.00
	theme–NOM only[1]	Yenghuy-NOM hug-psv	9	6.29
	agent–DAT only[1]	Chelswu-DAT hug-psv	0	0.00
	theme only, no CM[1]	Yenghuy-~~NOM~~ hug-psv	0	0.00
	agent only, no CM[1]	Chelswu-~~DAT~~ hug-psv	0	0.00
	undetermined, no CM[1]	Chelswu-~~NOM~~ hug-psv	0	0.00
Ditransitive	recipient–DAT only[2]	Chelswu-DAT give	5	3.50
	SUM		143	100.00

Note. CM = case-marking. 1) does not involve canonicity as it is undeterminable with only one overt argument. Although 2) does not relate to a transitive event *per se* and does not count as a relevant pattern, we considered it here because DAT is often used to indicate a recipient in the active and thus a potential competitor of the agent–DAT pairing in the passive.

The other notable finding in this analysis involved case-marking: the degrees of associations between individual markers and the corresponding functions diverged. Whereas NOM and ACC were related strongly to the agent and the theme, respectively, DAT was not likely to occur with the agent. Overall, the agent–NOM and theme–ACC pairings operated reliably, with the individual forms supplying the corresponding functions and vice versa. Of the two possible candidates of functions for NOM – the agent (in the active) and the theme (in the passive) – the former was predominant. Specifically, despite the positive values of ΔP scores involving ACC, this marker exhibited only a moderate level of association strength relative to the case of NOM, with ACC being more favourable as a cue to invite the theme than as an outcome to be invited by the theme.

On a related note, the strong bi-directional association between form and function that NOM manifests for transitive events supplies high cue validity, which increases cue strength enough to facilitate a learner's acquisition of this particular mapping early on (Bates & MacWhinney, 1982, 1989; MacWhinney, 1987). This language-specific feature in Korean seems somewhat inconsistent with the meaning-before-form account in language learning (Ramscar et al., 2010) and possibly serves as the core motivation for the early, rapid learning of this knowledge compared to ACC, as demonstrated in previous research (e.g., Cho, 1982; Jin et al., 2015; Kim et al., 2017; Lee, 2004; Lee & Cho, 2009; Lee et al., 2013; Shin, 2021).

Based on these results, we model children's knowledge about clause-level constructions through a Bayesian simulation, specifically asking whether and how the model learns the constructions in their entirety (i.e., without the mediation of lexical information). The findings of the caregiver input serve as a seed for the simulation, which models a learner's cognitive space regarding the two construction types in Korean. Our Bayesian learner acquires these constructions as schematised input, which comprises pairings of morpho–syntactic and semantic–functional properties representing these constructions.

Bayesian simulation

Bayesian inference assumes that humans continuously update their beliefs about an event, represented as probabilities, through accumulated observations, making inferences according to these updated beliefs. One's degree of belief about an event (posterior probability) is calculated using both the accumulated degree of conviction in a hypothesis which occurs before encountering the event (prior probability) and a conditional probability where the event would be observed given that the hypothesis is true (likelihood) (Pearl & Russell, 2001; Perfors et al., 2011a). This idea is formalised as Bayes' theorem (5), where A and B are independent events, $P(A|B)$ refers to the posterior probability, $P(B|A)$ to the likelihood, $P(A)$ to the prior probability, and $P(B)$ to the marginal probability.

$$P(A|B) = \frac{P(B|A)*P(A)}{P(B)} \qquad (5)$$

$P(B)$ is less important in actual application than in theory because the event B is fixed due to a stronger focus on the effects of the event A on one's beliefs (Kruschke, 2015). This condition produces a simpler formula (6) where the posterior probability is proportional to the likelihood times the prior probability (the marginal probability is not considered in this calculation).

$$P(A|B) \propto P(B|A)*P(A) \qquad (6)$$

Bayesian inference can accommodate how language develops with respect to lexico–grammatical knowledge (e.g., Alishahi & Stevenson, 2008; Bannard et al., 2009; Matusevych et al., 2016; Nguyen & Pearl, 2019; Xu & Tenenbaum, 2007; Perfors, Tenenbaum & Regier, 2011b), sentence-pattern-wise networks and productivity (e.g., Barak & Goldberg, 2017), and typological generalisations (e.g., Culbertson & Smolensky, 2012).

Alishahi and Stevenson (2008), *inter alia*, provided an important precedent for the current work. They showed a Bayesian account of the emergence and growth of English verb-argument constructions, which largely resembled developmental aspects that English-speaking children manifest. They created artificial input as pairs of a sentential frame and the corresponding semantic description involving this frame based on naturalistic caregiver input in CHILDES. These form–meaning pairs were inputted to an unsupervised Bayesian learning model to measure how the model displayed probability distributions in the formation of constructional clusters as learning proceeded. As the quantity of input increased over time, the Bayesian model was able to assign higher probabilities to frequently occurring verbs within specific constructions to which they were mapped and to generalise this schematic knowledge to a newly attested lexicon. Their modelling work is consistent with the major assumptions of the usage-based constructionist approach, supporting the interplay of frequency effects and general learning mechanisms without positing domain-specificity in language development.

Two conceptual points of Alishahi and Stevenson (2008) are highly relevant to our simulation. One is the direct mapping of a sentential frame and its semantic description. This reflects the idea that the inseparability of form and meaning/function, conceptualised as a construction, is a core property of language (Goldberg, 1995). We thus created input for this study's Bayesian learner by combining a constructional frame (a morpho–syntactic layer) and its meaning/function (a semantic–functional layer). The other point involves how constructions exist in humans' conceptual space. Alishahi and Stevenson (2008) assumed that constructional knowledge creates clusters that share similar features in their syntactic–semantic properties, intertwined with probabilities about how likely these clusters accord with or deviate from each other (cf. Goldberg, 2019). Following this point, we showed the development of constructional patterns (as clusters in the simulation environment) via posterior probabilities of these patterns and their changes due to learning.

Methods[8]

Composition of input

All the constructional patterns for transitive events were included, with scrambling and varying degrees of omission manifested (see Table 3). There is no Korean corpus of caregiver input paired with semantic–functional information, so we generated an artificial set of input based on the characteristics of Korean caregiver input in CHILDES pertaining to these patterns (cf. Alishahi & Stevenson, 2008). To focus exclusively on the development of constructional knowledge in its entirety, independently of lexical items, we devised two layers of schematised input: a morpho–syntactic layer specifying the formal properties of the pattern and a semantic–functional layer indicating the thematic roles of arguments and functions of markers. Each element in these layers had a left-to-right index to maintain information about canonicity in the input. For instance, the

---

[8]See this *github* repository for the simulation.

canonical active transitive (7) started with a nominal (N) followed by -*i/ka*, which was linked to the agent–nominative pair. This proceeded with another nominal followed by -*(l)ul*, which was associated with the theme–accusative pair, and finally a verb (V) denoting an action.

(7)   Example of input: canonical active transitive
      Morpho–syntactic layer   N_1–*i/ka*_1   N_2– *(l)ul*_2   V_3
      Semantic–functional layer Agent_1–NOM_1   Theme_2–ACC_2   Action_3

Whereas real morphemes indicated markers and passive morphology,[9] N and V represented abstract syntactic categories for nouns and verbs, respectively. Here, we did not presume that children receive these abstract categories from the beginning of learning; rather, we assumed that these categories represent HEURISTICS – strategic and provisional knowledge emerging probabilistically through exposure – employed during acquisition. That is, a word with a marker stands for an entity, and a word at the end of a sentence refers to an action. Notably, we included no content word to control for the effect of lexical information on the simulation results and to better demonstrate the developmental aspects of the constructional patterns in their entirety in the cognitive space that we modelled.

Model training

The general learning algorithm for our Bayesian learner was similar to that of Alishahi and Stevenson (2008): adding a new input item to an existing group of constructions that had the most similar characteristics to the item. The degree of similarity was determined by the probability that the new item was close to the individual groups of constructions existing in the model. This process is formalised as (8): to find the best-matching construction, the model classified a new input item *nCx* as an existing construction type *eCx*, ranging over the indices of all the constructions in the model, with the maximum probability given *nCx*.

$$\text{Best Construction}(nCx) = \underset{eCx}{\text{argmax}}\ P(eCx|nCx) \tag{8}$$

The computation of $P(eCx \mid nCx)$ followed Bayes' rule as in (6) where the posterior probability $P(eCx \mid nCx)$ was proportional to the multiplication of the conditional probabilities associated with the existing construction types and the priors of the existing construction types.

The actual frequency information in Table 3 served as initial priors for the constructional patterns. As learning proceeded, information about the constructional patterns was updated. This was achieved through updating the pattern frequencies by adding the number(s) of the classified input to the classified construction type over the course of

---

[9]In creating input, we did not consider allomorphy involving case-marking and passive morphology, assuming that the occurrence of allomorphy is evenly distributed. We acknowledge the possibility that one allomorph occurs more frequently than the others or that the degree of form–function mapping of individual allomorphs may be disproportionate. This remains as one limitation of this simulation work.
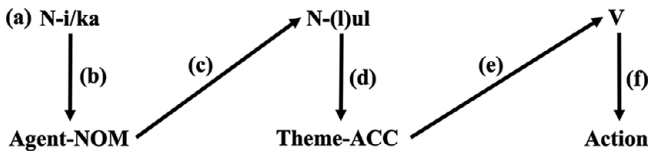
Figure 1. Schematic display of how to calculate transitional probability: canonical active transitive with no omission.

learning. To prevent the probability from converging upon zero, we adopted the Laplace smoothing technique (e.g., Agresti & Coull, 1998): the Laplace estimator added the value of one as the Laplace value to the original frequency value so that the probability of occurrence of each construction type did not become zero and thus incalculable.

For construction learning, we used transitional probability – namely, a series of conditional probabilities from the first item to the last item within a specific pattern. This reflects how children utilise linguistic input for learning – deducing intended meanings and functions from a given form (cf. Goldberg, 2019) – in an incremental manner (e.g., Özge, Küntay & Snedeker, 2019; Strotseva-Feinschmidt, Schipke, Gunter, Brauer & Friederici, 2019). To illustrate, the transitional probability of the canonical active transitive with no omission is obtained by the multiplication of the following probabilities (Figure 1): construction-initial N–*i/ka* pairing (a), construction-initial agent–NOM pairing given the construction-initial N–*i/ka* pairing (b), construction-medial N–*(l)ul* pairing given the construction-initial agent–NOM pairings (c), construction-medial theme–ACC pairing given the construction-medial N–*(l)ul* pairings (d), construction-final V given the construction-medial theme–ACC pairings (e), and construction-final action given the construction-final V (f). This particular composition nicely captures both pattern-wise facts (i.e., 'What items appear where and in what sequence?') and case-marking facts (i.e., 'What form–function associations of markers engage in a constructional pattern?') pertaining to a construction.

Model performance and prediction

There were 10 learning phases, with each phase consisting of one pass through the whole set of input (6,902 instances; see Table 3). Posterior probabilities of the constructional patterns were measured at every learning phase to estimate the degree of clustering for these constructions after the learning finished. We also traced the individual posterior probabilities from the learning phases 1–10 to see how the degree of clustering changed during learning in the given simulation environment.

We predicted two specific outcomes. First, the degree of clustering for the constructional patterns should be asymmetric as learning proceeds. The corpus analysis showed that factors such as thematic role ordering, the number of arguments, voice type, and the omission of sentential elements generate the construction asymmetry in the caregiver input, thereby manipulating the cue strength involving these patterns. This would create by-construction competition that should affect the course of learning. We thus expected a major increase in the posterior probabilities of the frequently attested patterns in the caregiver input (e.g., the canonical active transitive with no omission, the active transitive with only the theme–ACC pairing, the active transitive with only the theme argument without ACC). Furthermore, due to the characteristics of the Bayesian inference algorithm that constantly updates available information against previous experience, we

further anticipated a continuous increase in the posterior probabilities of these major patterns as the learning proceeded.

We also predicted that the growth of clustering for the suffixal passive patterns should be suppressed considerably throughout the learning process. We identified two core factors contributing to this suppression. At the construction level, there was an unusual occurrence of verbal morphology: compared to its active counterpart – the null (and default) form in the input – this construction type engaged in the passive suffix (PSV), which was scarce in the input. At the case-marking level, there were atypical form–function associations of case-marking: NOM indicating the theme (but typically the agent) and DAT indicating the agent (but typically the recipient), all of which were rare in the input. Therefore, we expected that the information from these two levels, together with the continuous updating mechanism in the Bayesian model, would inhibit the passive patterns across the board.

Results and discussion

Figure 2 presents posterior probabilities of the constructional patterns per learning. Whereas most of the constructional patterns converged upon almost zero probability, the canonical active transitive was the only pattern whose degree of clustering was constantly increasing as learning proceeded. This finding indicates that, because of both its high construction frequency and the typical or dominant type of form-function pairings of case-marking, this constructional pattern was well-established in our model. This also aligns with the findings of behavioural studies showing children's adult-like success in comprehending this pattern relative to the other patterns with a partial argument, marker, or both (e.g., Shin, 2021).

In contrast, the growth of several patterns did not comply with distributional properties in the input. The active transitive with only the theme–ACC pairing, for example, was the most frequent pattern in the given input (1,938 instances), but the posterior probability of this pattern was neither the highest nor did it defeat that of the canonical active transitive with no omission. The posterior probability of the active transitive with only the no-ACC theme argument, the third most frequent pattern in the input (1,155 instances), slightly increased until the fifth learning phase, when it then immediately decreased. One possible reason for this disparity is that the development of the clustering for these patterns was somehow inhibited due to the characteristics of the other active transitive patterns during learning. The initial theme–ACC pairing (1,989 instances) was outnumbered by the initial agent–NOM pairing (2,960 instances); the number of the initial no-ACC theme argument (1,161 instances) was smaller than that of the initial theme–ACC pairing (1,989 instances) and the initial agent–NOM pairing (2,960 instances). Therefore, this study's model may have learnt these case-marking properties (together with where each pairing occurred in a pattern) early and cumulatively during the learning process.

The degree of clustering for the remaining patterns decreased during the learning process, which may be ascribable to the same kind of suppression effects induced by these patterns' fully-fledged counterpart – the canonical active transitive with no omission, which occupied a fairly large amount of input. Meanwhile, the finding that the posterior probability of the active transitive with only the agent–NOM pairing decreased over learning remains unclear at this point. We speculate that a similar kind of inhibitory force, caused by various constructional patterns in the input, may have affected how this pattern
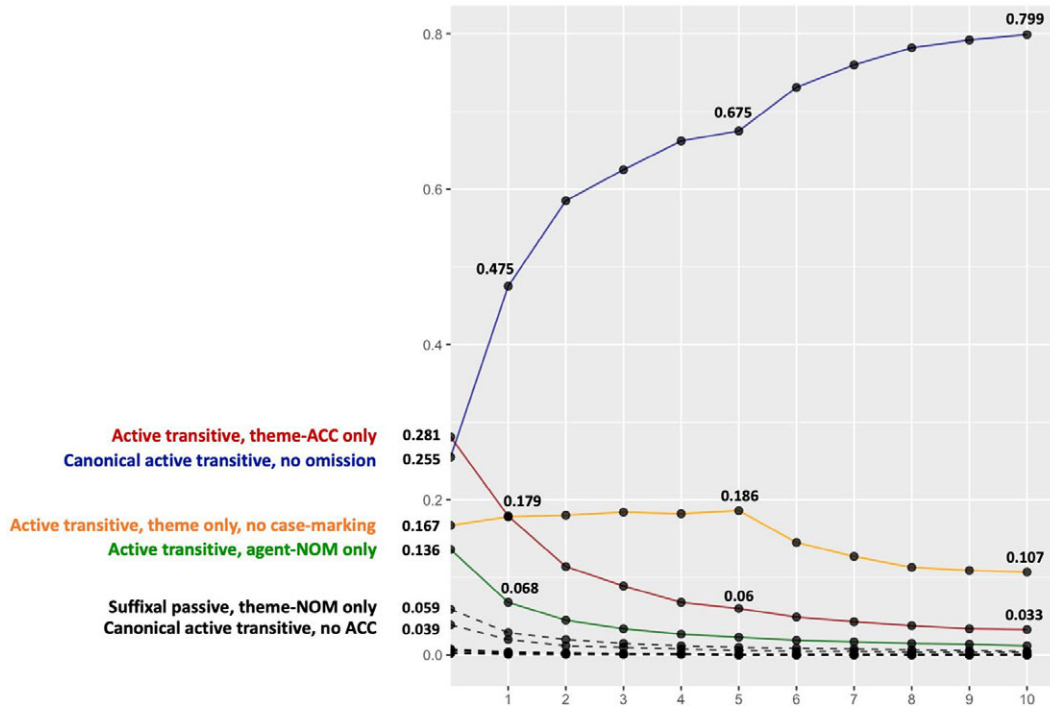
Figure 2. By-construction posterior probability per learning. X-axis: learning phase; Y-axis: posterior probability.
Note. The other constructional patterns not specified in this figure converged upon zero probability immediately after the first learning. The ditransitive pattern only with the recipient– DAT pairing did not fall into a transitive event and was thus excluded. For the readers' sake, this pattern achieved the posterior probability of 0.035 and 0.036 after the first and 10th learning, respectively.

was learnt. This pattern was the fourth most frequent one in the input. However, the agent–NOM pairing occurred less often before a verb (935 instances for the active transitive, agent–NOM only; six instances for the scrambled active transitive, no ACC) than before the N–*(l)ul* pairing (1,938 instances for the canonical active transitive, no omission). This interplay may have suppressed the development of this pattern despite its relatively high construction frequency and the typicality of the form–function mapping of case-marking in the pattern. Even so, the reason for this suppression remains unclear and may thus require further investigation.

The change of posterior probabilities in the passive patterns (the suffixal passive with only the theme–NOM pairing, only the agent–DAT pairing, or only the no-NOM theme argument) is attributable to cue competition involving case-marking and verbal morphology. The suffixal passive with only the theme–NOM pairing has two features: the unusual case-marking (i.e., NOM indicating the theme) and the atypical passive morphology. The development of this pattern may thus have been suppressed greatly by its corresponding pattern – the active transitive with only the agent–NOM pairing, which has the typical case-marking (NOM indicating the agent) and verbal morphology (no active morphology). Similarly, the growth of the suffixal passive with only the agent–DAT pairing may have been constrained by the ditransitive with only the recipient–DAT pairing: case-marking (DAT indicating the recipient is more often than DAT indicating the agent) and verbal morphology (verb with no morphology is more often than verb with passive suffixes). The suffixal passive with only the no-NOM theme argument engages in passive morphology, which is atypical; this nature may have facilitated a similar composition of this pattern, the active transitive patterns with only one case-less argument (1,248 instances), in expressing transitive events. These findings thus indicate that cue competition involving case-marking and verbal morphology across the two voice types substantially modulates learning outcomes in the model.

General discussion

Summary of findings

This study explored Korean-speaking children's knowledge about clause-level constructions in expressing a transitive event – an active transitive and a suffixal passive – in two ways: corpus analysis of caregiver input and child production (Section 3) and computational modelling through schematised input with no lexical information involved (Section 4).

The analysis of child corpora in CHILDES revealed two major aspects of caregiver input and child production. First, the rates of constructional patterns produced by the children generally mirrored those uttered by the caregivers. This aligns with the previous corpus-based studies across languages showing direct input–output relations in child language development (e.g., Ambridge et al., 2015b; Behrens, 2006; Cameron-Faulkner et al., 2003; Stoll et al., 2009). Second, despite the multiple form–function associations of case-marking in Korean, the caregivers' use of the three markers – NOM, ACC, and DAT – for the two construction types expressing transitive events was skewed towards single form–function pairings: NOM for the agent (and not for the theme), ACC for the theme (with uneven degrees of association between form and function by direction), and DAT not for the agent. These aspects provide empirical evidence for the nature of early input pertaining to the form–function mapping of case-marking dedicated to clause-level

constructions related to transitive events in Korean, which has remained unclear in the previous literature on Korean-speaking children's language development.

Based on the properties of the caregiver input, we modelled a Bayesian learner to see how the constructional patterns would develop as a result of the characteristics of construction-based input (without considering lexical information), by measuring the patterns' posterior probabilities over the course of learning. Overall, we found the dominance of one pattern, the canonical active transitive with no omission, which occupied approximately one-third of both the caregiver input and the children's production regarding constructional patterns expressing transitive events. In contrast, the development of the other patterns, including the one-argument active pattern either with only the theme–ACC pairing or with only the no-ACC theme argument and the passive patterns, seemed to be suppressed. The disproportionate rate of learning outcomes suggests that input properties, together with a statistical learning mechanism, may shape the structure of linguistic knowledge in a way that drives such information to centre around the most representative frame. Together, the simulation results suggest that this study's learning model could reveal reasonable linguistic generalisations, by forming constructional knowledge as a function of schematised input and statistical learning, even in the case of lesser-studied languages such as Korean. We believe that the particular information that we utilised for the model training – transitional probability – allowed the model to achieve this degree of generalisation, by incorporating the constructional distributions and the particular form–function mapping of the core structural components of each construction type.

Inconsistency in the development of constructional patterns across corpus analysis and Bayesian simulation

This global similarity between the model performance and the children's production is tempered by some notable inconsistencies, which are summarised in Table 7 (see also Appendix for the whole comparison between the caregiver input, children's production, and posterior probabilities of the constructional patterns at the 10th learning phase). Considering the overall number of constructional patterns that the children produced (143 instances), they seemed to prefer the three patterns, all of which include NOM, in production. In contrast, the learning model did not yield the corresponding rates of posterior probabilities for these patterns within the given simulation environment.

Table 7. Three constructional patterns involving major inconsistencies across corpus analysis and simulation (10th learning).

| | Corpus analysis | | | | |
| | Caregiver input | | Child production | | |
Construction	#	%	#	%	Simulation (posterior probability)
Active transitive, agent–NOM only	935	13.55	21	14.69	0.012
Canonical active transitive, no ACC	268	3.88	14	9.79	0.004
Suffixal passive, theme–NOM only	407	5.90	9	6.29	0.005

It seems that our computational model faithfully followed the construction-based distributional properties attested in the caregiver input. For instance, the active transitive with only the agent–NOM pairing (935 instances) was outnumbered by the corresponding pattern with only the theme–ACC pairing (1,938 instances), which may have affected the posterior probability of the former pattern through the raw frequency. The canonical active transitive with no ACC (268 instances) also occurred less frequently than its fully equipped counterpart (canonical active transitive with no omission: 1,757 instances). This may have influenced the posterior probability of this pattern through both the raw frequency and the transitional probability – that is, P(*Theme_2–ACC_2* | *Agent_1–NOM_1*) suppressed P(*N_2* | *Agent_1–NOM_1*). In the same way, the number of the suffixal passive with only the theme–NOM pairing (407 instances) was less than that of the active transitive with only the agent–NOM pairing (935 instances), and this may have guided the posterior probability of the passive pattern by way of both the raw frequency and the transitional probability – that is, P(*Agent_1–NOM_1* | *N_1-i/ka_1*) suppressed P(*Theme_1–NOM_1* | *N_1-i/ka_1*). Considering that our learning model proceeded with transitional probabilities accounting for both constructional distributions and case-marking facts, it is reasonable to think that the model responded favourably to the construction frequency and the form–function mapping of case-marking in the input.

The children, however, may have been affected more by the reliable or available form–function mapping of NOM for transitive events than constructional distributions in the caregiver input. We found in the corpus analysis that (i) NOM was not only a highly reliable cue to introduce the agent but also a highly reliable outcome invited by the agent and (ii) it occurred more frequently in the initial position than in the non-initial position. In turn, these characteristics allow for high cue validity for this particular mapping (Bates & MacWhinney, 1989), leading the children to primarily (and strongly) deploy NOM to indicate the actor of a transitive event. This interpretation supports previous research demonstrating the Korean-speaking children's heavy reliance on a heuristic that maps NOM onto the agent role (particularly for the first noun) for transitive constructions (e.g., Jin et al., 2015; Kim et al., 2017; Lee et al., 2013; Shin, 2021). Indeed, children are known to be better attuned to a local cue (induced by case-marking) than to a distributional cue (induced by word order) due to the computational advantage of the former versus the latter (Bates & MacWhinney, 1989; Shin, 2021; Wittek & Tomasello, 2005). In this respect, compared to the computational model that considers both local and distributional cues simultaneously, the children may have attended more to the agent–NOM pairing than the construction-based distributional properties in the early stages of learning.

Nevertheless, the case of the suffixal passive with only the theme–NOM pairing is still unclear. We speculate that there was some influence of lexical items on this inconsistency. As reported in the corpus analysis, the way that the children produced this pattern was tied to several verbs. Despite the numeric insufficiency for generalisation, it seems that the children's production of this pattern was limited to less abstract, narrow-range schemata, which is consistent with the gradual-abstraction account (e.g., Tomasello, 2003). This lexical specificity found in the passive may be due to the challenge of learning a passive voice (e.g., Kim et al., 2017; Shin, 2020; also cross-linguistically e.g., Abbot-Smith et al., 2017; Huang et al., 2013). Even so, because no content word was used in the input for the present simulation, this issue is left unaddressed in the current study and requires further investigation.

Broader implications on child language development

Our simulation work provided somewhat different flavour than the previous research on this subject (e.g., Alishahi & Stevenson, 2008; Ambridge & Blything, 2016; Bannard et al., 2009; Barak et al., 2016; Matusevych et al., 2016) due to the two motivations of this study. One was to model a child learner after the age of one or two, following the age range of the children in CHILDES (see Table 1). For this reason, we employed frequency information in the caregiver input as the initial priors of the learning model, instead of creating a *tabula rasa* model from scratch, with the assumption that this study's Bayesian learner was already equipped with varying degrees of prior probabilities involving the constructional patterns.[10] The other motivation in this study was to model the development of linguistic knowledge about clause-level constructions in their entirety. For this reason, we devised the schematised input with a pair of two abstract layers, instead of using content words attested in the caregiver input. Therefore, this study's computational model cannot predict whether children's linguistic knowledge is organised around specific lexical items and develops towards abstract constructions in a piecemeal manner, as the gradual abstraction account claims (e.g., Theakston, Ibbotson, Freudenthal, Lieven & Tomasello, 2015; Tomasello, 1992, 2003). Instead, this particular simulation environment allowed us to test how the Bayesian model learns constructional knowledge as proposed by the early abstraction account, the other perspective of the usage-based constructionist approach arguing for the early emergence of abstract knowledge (albeit still requiring a considerable amount of exposure to linguistic environments for the maturation of knowledge; e.g., Dąbrowska & Tomasello, 2008; Rowland, Chang, Ambridge, Pine & Lieven, 2012; Saffran, Aslin & Newport, 1996; cf. Messenger & Fisher, 2018).

Due to these motivations and the particularities for the simulation environment, in conjunction with this study's narrow scope of investigation (i.e., constructions only for transitive events), our computational model may not have exactly demonstrated human linguistic behaviours, as shown in the children's production. In particular, the fact that we composed input without lexical information renders it impossible for the model to capture this lexically tied factor (cf. Alishahi & Stevenson, 2008) to the extent that human learners do when acquiring constructional knowledge (e.g., Ambridge, Bidgood, Twomey, Pine, Rowland & Freudenthal, 2015a; Goldberg, 2019; Tomasello, 2003), as in the case of the suffixal passive with only the theme–NOM pairing. Furthermore, we utilised only well-formed instances (with at least one argument and a verb), ignoring incomplete instances in the caregiver input such as partial and verb-less utterances with various noun–marker pairings.[11] Therefore, the answer to the core question of this study can only be partial at this point.

---

[10]One reviewer suggested a comparison between the current model with predetermined priors and another model with uninformative priors. Although we agree with the value of this suggestion, we are somewhat hesitant with conducting this comparison in the present study for the following reasons. First, the uninformative-priors model does not reflect the core motivation of our modelling study – modelling a child learner after the age of one or two (which is the age range of the children in CHILDES). Second, the development of the uninformative-priors model would prove difficult. To devise this model, we must determine the extent to which priors are uninformative (or objective); unfortunately, we are not aware of previously set standards for determining uninformative priors that would be appropriate for our model architecture. These reasons make the development of an uninformative-priors model beyond the scope of the current study. We leave this issue open for now, and we believe this work to merit further inquiry.

[11]One reviewer suggested running the model with input in which some portions are truncated in some way. We believe that testing the impact of partial utterances on model performance is necessary and desirable, but we think that its preparation/implementation would be difficult. To conduct this work, the number of

Nonetheless, we discovered convincing compatibility of the model performance with the children's production. For instance, the distributional properties of the constructional patterns for transitive events and the characteristics of case-marking and verbal morphology dedicated to these constructions in the caregiver input yielded the model performance largely consistent with the children's production, despite having no individual support from lexical information. This approximates how Korean-speaking children's constructional knowledge develops and changes in their conceptual space in response to construction frequency within the given amount of input and form–function correlations involving the core structural properties of the target construction types. In particular, the suppression effects observed in the model performance reflects the by-construction competition, driven by the asymmetric degrees of cue validity induced by both constructions and their structural components (i.e., case-marking and verbal morphology). This aligns nicely with the Competition Model that shows how children acquire coalitions of form–function mapping and adjust the weight of each mapping for an optimal fit for learning (Bates & MacWhinney, 1982, 1989; MacWhinney, 1987).

Our findings also highlight the status of abstract form–function correspondences – constructions, which are independent of individual lexical items – as a psychological reality in language development (Goldberg, 2019; Lieven, 2016; Tomasello, 2009). The classic version of computational simulations within the usage-based constructionist approach has been utilising both lexically specific information and constructional information simultaneously (e.g., Alishahi & Stevenson, 2008; Ambridge & Blything, 2016; Bannard et al., 2009; Barak et al., 2016; Barak & Goldberg, 2017). However, this study exclusively considered information about constructions, with a special focus on the distributional properties of constructional patterns and the particular form–function pairings of the core structural components for each construction type. This aspect may render it somewhat difficult to pinpoint the locus of the dissimilarities between the caregiver input, the children's production, and the model performance. However, this study's novel approach allows us to effectively examine the extent to which children respond to knowledge about clause-level constructions during learning.

Together, the present study contributes to the literature on child language development in two directions. First, the implications of our findings support the major tenet of the usage-based constructionist approach that explains the development of linguistic knowledge as a result of the interplay between input properties and domain-general learning capacities (Ambridge et al., 2015b; Goldberg, 2019; Lieven, 2010; Tomasello, 2003). Second, this study's implications expand the current research practice in computational modelling for child language to include the unit of clause-level construction (without the mediation of lexical information). Specifically, this study employed direct form–function mapping and transitional probability for model training, illuminating the role of core morpho–syntactic features comprising the target construction types (case-marking and verbal morphology; scrambling or omission of sentential components) in the model's construction learning. In conclusion, we believe this study's findings advance understanding of how input-related factors (the nature of item frequency/distribution

---

partial utterances relevant to the current input that exist in the corpora must be identified. This must proceed manually, which is extremely challenging considering the entire input size. We might add some arbitrary numbers of partial utterance to the existing input, but there is no theoretical/empirical support for justifying arbitrary values. We are thus hesitant with this extension in the present study. Future studies should seek to clarify the degree to which partial and/or verb-less utterances (particularly involving noun–marker pairings) in caregiver input contribute to model performance.

and form–function associations) and learning mechanisms (statistical learning, together with the continuously updating mechanism against prior experience, as Bayesian inference suggests) jointly affect the organisation of target linguistic knowledge (clause-level construction) in children's cognitive space – particularly, regarding lesser-studied languages in this field.

The findings of this study should be further verified and re-assessed from various angles, particularly through behavioural experiments on (the structural components of) clause-level constructions. Compared to the active employment of the real-time measurement of children's sentence comprehension in major languages under investigation (e.g., Abbot-Smith et al., 2017; Huang et al., 2013; Özge et al., 2019; Strotseva-Feinschmidt et al., 2019), processing-based research on child language in Korean is in its infancy. Furthermore, the literature is scant on Korean-speaking children's linguistic development considering language-specific properties at the level of clause-level constructions (cf. Jin et al., 2015; Kim et al., 2017; Lee et al., 2013). With a similar focus on the two construction types in this study, Shin (2020) revealed an interplay of word order, case-marking, and verbal morphology in Korean-speaking children's comprehension of these constructions with scrambling and varying degrees of omission. By devising a novel methodology that obscured parts of test sentences through acoustic masking, Shin found a comprehension advantage of a local cue (case-marking; particularly the agent–NOM pairing) over a distributional cue (word order; particularly the AGENT-FIRST heuristic) and emerging sensitivity to passive morphology proportionate to age. Future work would thus benefit from exploring to what extent the findings of computational simulations (with various learning algorithms) explain those from behavioural experiments. This is what we plan to pursue next.

Competing interests. The authors declare none.

References

Abbot-Smith, K., **Chang, F.**, **Rowland, C.**, **Ferguson, H.**, & **Pine, J.** (2017). Do two and three year old children use an incremental first-NP-as-agent bias to process active transitive and passive sentences?: A permutation analysis. *PloS one*, **12**(10), e0186129. https://doi.org/10.1371/journal.pone.0186129

Abbot-Smith, K., & **Tomasello, M.** (2006). Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *The Linguistic Review*, **23**(3), 275–290. https://doi.org/10.1515/TLR.2006.011

Agresti, A., & **Coull, B. A.** (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician*, **52**(2), 119–126. https://doi.org/10.1080/00031305.1998.10480550

Aguado-Orea, J., & **Pine, J. M.** (2015). Comparing different models of the development of verb inflection in early child Spanish. *PloS one*, **10**(3), e0119613. https://doi.org/10.1371/journal.pone.0119613

Alishahi, A., & **Stevenson, S.** (2008). A computational model of early argument structure acquisition. *Cognitive Science*, **32**(5), 789–834. https://doi.org/10.1080/03640210801929287

Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society*, **15**(3), 147–149. https://doi.org/10.3758/BF03334492

Ambridge, B., **Bidgood, A.**, **Twomey, K. E.**, **Pine, J. M.**, **Rowland, C. F.**, & **Freudenthal, D.** (2015a). Preemption versus entrenchment: Towards a construction-general solution to the problem of the retreat from verb argument structure overgeneralization. *PloS one*, **10**(4), e0123723. https://doi.org/10.1371/journal.pone.0123723

Ambridge, B., & **Blything, R. P.** (2016). A connectionist model of the retreat from verb argument structure overgeneralization. *Journal of Child Language*, **43**(6), 1245–1276. https://doi.org/10.1017/S0305000915000586

Ambridge, B., **Kidd, E.**, **Rowland, C. F.**, & **Theakston, A. L.** (2015b). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, **42**(2), 239–273. https://doi.org/10.1017/S030500091400049X

Ambridge, B., **Maitreyee, R.**, **Tatsumi, T.**, **Doherty, L.**, **Zicherman, S.**, **Pedro, P. M.**, **Bannard, C.**, **Samanta, S.**, **McCauley, S.**, **Arnon, I.**, **Bekman, D.**, **Efrati, A.**, **Berman, R.**, **Narasimhan, B.**, **Sharma, D. M.**, **Nair, R. B.**, **Fukumura, K.**, **Campbell, S.**, **Pye, C.**, **Pixabaj, S. F. C.**, **Paliz, M. M.**, & **Mendoza, M. J.** (2020). The crosslinguistic acquisition of sentence structure: Computational modeling and grammaticality judgments from adult and child speakers of English, Japanese, Hindi, Hebrew and K'iche'. *Cognition*, **202**, 104310. https://doi.org/10.1016/j.cognition.2020.104310

Bannard, C., **Lieven, E.**, & **Tomasello, M.** (2009). Modeling children's early grammatical knowledge. *Proceedings of the National Academy of Sciences*, **106**(41), 17284–17289. https://doi.org/10.1073/pnas.0905638106

Barak, L., **Goldberg, A. E.**, & **Stevenson, S.** (2016). Comparing computational cognitive models of generalization in a language acquisition task. In **J. Su** & **K. Duh** & **X. Carreras** (Eds.), *Proceedings of the 2016 conference on Empirical Methods in Natural Language Processing* (pp. 96–106). Association for Computational Linguistics.

Barak, L., & **Goldberg, A.** (2017). Modeling the partial productivity of constructions. In *the proceeding of the Association for the Advancement of Artificial Intelligence 2017 Spring Symposium on Computational Construction Grammar and Natural Language Understanding: Technical Report SS-17-02* (pp.131–138). AAAI.

Bates, E., & **MacWhinney, B.** (1982). Functionalist approaches to grammar. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 173–218). New York, NY: Cambridge University Press.

Bates, E., & **MacWhinney, B.** (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The cross-linguistic study of sentence processing* (pp. 3–76). New York, NY: Cambridge University Press.

Behrens, H. (2006). The input–output relationship in first language acquisition. *Language and Cognitive Processes*, **21**(1-3), 2–24. https://doi.org/10.1080/01690960400001721

Cameron-Faulkner, T., **Lieven, E.**, & **Tomasello, M.** (2003). A construction based analysis of child directed speech. *Cognitive Science*, **27**(6), 843–873. https://doi.org/10.1207/s15516709cog2706_2

Cameron-Faulkner, T., **Lieven, E.**, & **Theakston, A.** (2007). What part of no do children not understand? A usage-based account of multiword negation. *Journal of Child Language*, **34**(2), 251–282. https://doi.org/10.1017/S0305000906007884

Cho, S. W. (1982). *The acquisition of word order in Korean.* Unpublished Master's thesis. Department of Linguistics, University of Calgary.

Choo, M., & **Kwak, H-Y.** (2008). *Using Korean*. Cambridge: Cambridge University Press.

Chung, G. (1994). *Case and its acquisition in Korean.* Unpublished Ph.D. dissertation. Department of Linguistics, University of Texas at Austin.

Culbertson, J., & **Smolensky, P.** (2012). A Bayesian model of biases in artificial language learning: The case of a word-order universal. *Cognitive Science*, **36**(8), 1468–1498. https://doi.org/10.1111/j.1551-6709.2012.01264.x

Dąbrowska, E., & **Tomasello, M.** (2008). Rapid learning of an abstract language-specific category: Polish children's acquisition of the instrumental construction. *Journal of Child Language*, **35**(3), 533–558. https://doi.org/10.1017/S0305000908008660

Desagulier, G. (2016). A lesson from associative learning: asymmetry and productivity in multiple-slot constructions. *Corpus Linguistics and Linguistic Theory*, **12**(2), 173–219. https://doi.org/10.1515/cllt-2015-0012

Dittmar, M., **Abbot-Smith, K.**, **Lieven, E.**, & **Tomasello, M.** (2008). German children's comprehension of word order and case marking in causative sentences. *Child Development*, **79**(4), 1152–1167. https://doi.org/10.1111/j.1467-8624.2008.01181.x

Garcia, R., & **Kidd, E.** (2020). The acquisition of the Tagalog symmetrical voice system: Evidence from structural priming. *Language Learning and Development*, **16**(4), 1–27. https://doi.org/10.1080/15475441.2020.1814780

Goldberg, A. E. (1995). *Constructions: a construction grammar approach to argument structure*. Chicago, IL: University of Chicago Press.

Goldberg, A. E. (2019). *Explain me this: Creativity, competition, and the partial productivity of constructions*. Princeton, NJ: Princeton University Press.

Huang, Y. T., Zheng, X., Meng, X., & Snedeker, J. (2013). Children's assignment of grammatical roles in the online processing of Mandarin passive sentences. *Journal of Memory and Language*, **69**(4), 589–606. https://doi.org/10.1016/j.jml.2013.08.002

Ibbotson, P., & Tomasello, M. (2009). Prototype constructions in early language acquisition. *Language and Cognition*, **1**(1), 59–85. https://doi.org/10.1515/LANGCOG.2009.004

Jin, K-S., Kim, M. J., & Song, H-J. (2015). The development of Korean preschooler' ability to understand transitive sentences using case-markers. *The Korean Journal of Cognitive and Biological Psychology*, **28**(3), 75–90.

Kim, M. (2010). Syntactic priming in children's production of passives. *Korean Journal of Applied Linguistics*, **26**(2), 271–290.

Kim, S., O'Grady, W., & Cho, S. (1995). The acquisition of case and word order in Korean. *Language Research*, **31**(4), 687–695.

Kim, S. Y., Sung, J. E., & Yim, D. (2017). Sentence comprehension ability and working memory capacity as a function of syntactic structure and canonicity in 5-and 6-year-old children. *Communication Sciences & Disorders*, **22**(4), 643–656. https://doi.org/10.12963/csd.17420

Kim, Y. J. (1997). The acquisition of Korean. In D. Slobin (Ed.), The Crosslinguistic Study of Language Acquisition 4 (pp. 335–443). Hillsdale, NJ: Lawrence Erlbaum.

Kruschke, J. (2015). *Doing Bayesian data analysis: A Tutorial with R, JAGS, and Stan* (2nd edition). London: Elsevier.

Langacker, R. W. (2017). Entrenchment in Cognitive Grammar. In H-J Schmid (Ed.), *Entrenchment and the psychology of language learning: how we reorganize and adapt linguistic knowledge* (pp. 39–56). Berlin: Walter de Gruyter.

Lee, C., & Cho, S. W. (2009). Acquisition of the subject and topic nominals and markers in the spontaneous speech of young children in Korean. In C. Lee, G. B. Simpson & Y. Kim (Eds.), *The Handbook of East Asian Psycholinguistics 3* (pp. 23–33). New York, NY: Cambridge University Press.

Lee, H. R. (2004). 2sey hankwuk atonguy cwue paltal thukseng [A study of early subject acquisition in Korean]. *Communication Sciences and Disorders*, **9**(2), 19–32.

Lee, K. O., & Lee, Y. (2008). An event-structural account of passive acquisition in Korean. *Language and Speech*, **51**(1/2), 133–149. https://doi.org/10.1177/00238309080510010801

Lee, Y. L., Kim, M. J., & Song, H. J. (2013). The development of Korean children's abilities to use structural cues for sentence comprehension. *The Korean Journal of Developmental Psychology*, **26**(4), 125–139.

Lieven, E. (2010). Input and first language acquisition: Evaluating the role of frequency. *Lingua*, **120**(11), 2546–2556. https://doi.org/10.1016/j.lingua.2010.06.005

Lieven, E. (2016). Usage-based approaches to language development: Where do we go from here?. *Language and Cognition*, **8**(3), 346–368. https://doi.org/10.1017/langcog.2016.16

Lupyan, G., & Christiansen, M. H. (2002, January). Case, word order, and language learnability: Insights from connectionist modeling. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 24. Retrieved from https://escholarship.org/uc/item/8nf95595 on 03 July 2019

MacWhinney, B. (1987). The Competition Model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 249–308). Hillsdale, NJ: Lawrence Erlbaum.

MacWhinney, B. (2000). The CHILDES project: Tools for analyzing talk (3rd edition). Mahwah, NJ: Lawrence Erlbaum.

Matusevych, Y., Alishahi, A., & Backus, A. (2016). Modelling verb selection within argument structure constructions. *Language, Cognition and Neuroscience*, **31**(10), 1215–1244. https://doi.org/10.1080/23273798.2016.1200732

Messenger, K., & Fisher, C. (2018). Mistakes weren't made: Three-year-olds' comprehension of novel-verb passives provides evidence for early abstract syntax. *Cognition*, **178**, 118–132. https://doi.org/10.1016/j.cognition.2018.05.002

Nguyen, E., & Pearl, L. (2019). Using developmental modeling to specify learning and representation of the passive in English children. In M. M. Brown & B. Dailey (Eds.), *Proceedings of the 43rd Boston University Conference on Language Development* (pp. 469–482). Somerville, MA: Cascadilla Press.

No, G. (2009). Acquisition of case markers and grammatical functions. In C. Lee, G. B. Simpson & Y. Kim (Eds.), *The Handbook of East Asian Psycholinguistics* (Vol. **3**) (pp. 51–62). Cambridge, UK: Cambridge University Press.

Özge, D., Küntay, A., & Snedeker, J. (2019). Why wait for the verb? Turkish speaking children use case markers for incremental language comprehension. *Cognition*, **183**, 152–180. https://doi.org/10.1016/j.cognition.2018.10.026

Pearl, J., & Russell, S. (2001). Bayesian networks. In M. A. Arbib (Ed.), *The handbook of brain theory and neural networks* (pp. 157–159). Boston, MA: MIT Press.

Perfors, A., Tenenbaum, J. B., Griffiths, T. L., & Xu, F. (2011a). A tutorial introduction to Bayesian models of cognitive development. *Cognition*, **120**(3), 302–321. https://doi.org/10.1016/j.cognition.2010.11.015

Perfors, A., Tenenbaum, J. B., & Regier, T. (2011b). The learnability of abstract syntactic principles. *Cognition*, **118**(3), 306–338. https://doi.org/10.1016/j.cognition.2010.11.001

Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science*, **34**, 909–957. https://doi.org/10.1111/j.1551-6709.2009.01092.x

Rowland, C. F., Chang, F., Ambridge, B., Pine, J. M., & Lieven, E. V. (2012). The development of abstract syntax: Evidence from structural priming and the lexical boost. *Cognition*, **125**(1), 49–63. https://doi.org/10.1016/j.cognition.2012.06.008

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, **274**(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Shin, G-H. (2020). *Connecting input to comprehension: First language acquisition of active transitives and suffixal passives by Korean-speaking preschool children.* Unpublished Ph.D. dissertation. University of Hawaiʻi at Mānoa.

Shin, G-H. (2021). Limits on the Agent-First strategy: Evidence from children's comprehension of a transitive construction in Korean. *Cognitive Science*, **45**(9), e13038. https://doi.org/10.1111/cogs.13038

Sohn, H. M. (1999). *The Korean language.* Cambridge University Press.

Stefanowitsch, A. (2011). Constructional preemption by contextual mismatch: A corpus-linguistic investigation. *Cognitive Linguistics*, **22**(1), 107–129. https://doi.org/10.1515/cogl.2011.005

Stoll, S., Abbot-Smith, K., & Lieven, E. (2009). Lexically restricted utterances in Russian, German, and English child-directed speech. *Cognitive Science*, **33**(1), 75–103. https://doi.org/10.1111/j.1551-6709.2008.01004.x

Strotseva-Feinschmidt, A., Schipke, C. S., Gunter, T. C., Brauer, J., & Friederici, A. D. (2019). Young children's sentence comprehension: Neural correlates of syntax-semantic competition. *Brain and Cognition*, **134**, 110–121. https://doi.org/10.1016/j.bandc.2018.09.003

Theakston, A. L. (2004). The role of entrenchment in children's and adults' performance on grammaticality judgment tasks. *Cognitive Development*, **19**(1), 15–34. https://doi.org/10.1016/j.cogdev.2003.08.001

Theakston, A. L., Ibbotson, P., Freudenthal, D., Lieven, E. V., & Tomasello, M. (2015). Productivity of noun slots in verb frames. *Cognitive Science*, **39**(6), 1369–1395. https://doi.org/10.1111/cogs.12216

Tomasello, M. (1992). *First verbs: A case study of early grammatical development.* New York, NY: Cambridge University Press.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition.* Cambridge, MA: Harvard University Press.

Tomasello, M. (2009). The usage-based theory of language acquisition. In E. L. Bavin (Ed.), *The Cambridge handbook of child language* (pp. 69–87). Cambridge: Cambridge University Press.

Wittek, A., & Tomasello, M. (2005). German-speaking children's productivity with syntactic constructions and case morphology: Local cues act locally. *First Language*, **25**(1), 103–125. https://doi.org/10.1177/0142723705049120

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, **114**(2), 245–272. https://doi.org/10.1037/0033-295X.114.2.245

Yeon, J. (2015). Passives. In L. Brown & J. Yeon (Eds.), *The handbook of Korean linguistics* (pp. 116–136). Oxford: John Wiley & Sons.

Appendix

Comparison between the caregiver input, children's production, and posterior probabilities (10th learning) of the constructional patterns.

		Corpus analysis				Simulation (posterior probability)
		Caregiver input		Child production		
	Construction	#	%	#	%	
Canonical active transitive	No omission	1,757	25.46	37	25.87	0.799
	no ACC	268	3.88	14	9.79	0.004
	no NOM	19	0.28	0	0.00	< 0.001
Scrambled active transitive	No omission	51	0.74	0	0.00	0.001
	no NOM	0	0.00	0	0.00	< 0.001
	no ACC	6	0.09	0	0.00	< 0.001
Active transitive with omission	agent–theme, no CM	3	0.04	0	0.00	< 0.001
	theme–agent, no CM	0	0.00	0	0.00	< 0.001
	undetermined, no CM	0	0.00	0	0.00	< 0.001
	agent–NOM only[1]	935	13.55	21	14.69	0.012
	theme–ACC only[1]	1,938	28.08	25	17.48	0.033
	agent only, no CM[1]	53	0.77	1	0.70	0.001
	theme only, no CM[1]	1,155	16.73	30	20.98	0.107
	undetermined, no CM[1]	40	0.58	1	0.70	< 0.001
Canonical suffixal passive	No omission	2	0.03	0	0.00	< 0.001
	no DAT	0	0.00	0	0.00	< 0.001
	no NOM	0	0.00	0	0.00	< 0.001
Scrambled suffixal passive	No omission	1	0.01	0	0.00	< 0.001
	no NOM	0	0.00	0	0.00	< 0.001
	no DAT	0	0.00	0	0.00	< 0.001
Suffixal passive with omission	theme–agent, no CM	0	0.00	0	0.00	< 0.001
	agent–theme, no CM	0	0.00	0	0.00	< 0.001
	undetermined, no CM	0	0.00	0	0.00	< 0.001
	theme–NOM only[1]	407	5.90	9	6.29	0.005
	agent–DAT only[1]	13	0.19	0	0.00	< 0.001
	theme only, no CM[1]	20	0.29	0	0.00	< 0.001
	agent only, no CM[1]	0	0.00	0	0.00	< 0.001
	undetermined, no CM[1]	0	0.00	0	0.00	< 0.001
Ditransitive	recipient–DAT only[2]	234	3.39	5	3.50	< 0.001
	SUM	6,902	100.00	143	100.00	1.000