Bayesian simulation of Korean-speaking children's constructional knowledge about transitive events

Gyu-Ho Shin (Palacky University Olomouc)

Seongmin Mun (Chosun University)

Bayesian inference

- Humans keep updating their beliefs about an event (represented as probabilities) through accumulated observations and make inferences by way of the updated beliefs

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

  - The degree of belief about an event (*posterior probability*) is calculated jointly by 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*)

Research question

- How do Korean-speaking children shape **clause-level constructional knowledge** as a function of exposure to language and domain-general learning capacities?
  - Target: knowledge about representative argument structure constructions involving a transitive event in Korean (*active transitives & suffixal passives*)
  - Approach: Bayesian simulation that employs information about the frequency of the two construction types found in CHILDES (the largest & open-access child language corpora)

Key study

- Alishahi and Stevenson (2008)
  - Demonstrated a Bayesian way of emergence and growth of English verb-argument constructions
  - Two conceptual points highly relevant to this study's aim
    - direct mapping of a sentential frame and its semantic description
      - the *inseparability* of form and meaning/function (conceptualised as a construction) is a core property of language
    - how constructions exist in humans' cognitive space
      - constructional knowledge creates *clusters* that share similar features in their syntactic-semantic properties, intertwined with *probabilities* about how likely they accord with or deviate from each other

Methods

- Input composition

Table 1. Constructional patterns (with or without scrambling and omission of sentential components) for a transitive event in CHILDES

Construction		Example	Freq (#) Caregiver	Child
Canonical active transitive	No omission	Mina-NOM Ciwu-ACC hug	1,757	37
	no ACC	Mina-NOM Ciwu-~~ACC~~ hug	268	14
	no NOM	Mina-~~NOM~~ Ciwu-ACC hug	19	0
Scrambled active transitive	No omission	Ciwu-ACC Mina-NOM hug	51	0
	no NOM	Ciwu-ACC Mina-~~NOM~~ hug	0	0
	no ACC	Ciwu-~~ACC~~ Mina-NOM hug	6	0
Active transitive	agent-theme, no CM	Mina-~~NOM~~ Ciwu-~~ACC~~ hug	3	0
	theme-agent, no CM	Ciwu-~~ACC~~ Mina-~~NOM~~ hug	0	0
	undetermined, no CM	Mina-~~NOM~~ Ciwu-~~ACC~~ hug	0	0
	agent-NOM only[1)]	Mina-NOM hug	935	21
	theme-ACC only[1)]	Ciwu-ACC hug	1,938	25
	agent only, no CM[1)]	Mina-~~NOM~~ hug	53	1
	theme only, no CM[1)]	Ciwu-~~ACC~~ hug	1,155	30
	undetermined, no CM[1)]	Mina-~~NOM~~ hug	40	1
Canonical suffixal passive	No omission	Ciwu-NOM Mina-DAT hug-psv	2	0
	no DAT	Ciwu-NOM Mina-~~DAT~~ hug-psv	0	0
	no NOM	Ciwu-~~NOM~~ Mina-DAT hug-psv	0	0
Scrambled suffixal passive	No omission	Mina-DAT Ciwu-NOM hug-psv	1	0
	no NOM	Mina-DAT Ciwu-~~NOM~~ hug-psv	0	0
	no DAT	Mina-~~DAT~~ Ciwu-NOM hug-psv	0	0
Suffixal passive	theme-agent, no CM	Ciwu-~~NOM~~ Mina-~~DAT~~ hug-psv	0	0
	agent-theme, no CM	Mina-~~DAT~~ Ciwu-~~NOM~~ hug-psv	0	0
	undetermined, no CM	Ciwu-~~NOM~~ Mina-~~ACC~~ hug-psv	0	0
	theme-NOM only[1)]	Ciwu-NOM hug-psv	407	9
	agent-DAT only[1)]	Mina-DAT hug-psv	13	0
	theme only, no CM[1)]	Ciwu-~~NOM~~ hug-psv	20	0
	agent only, no CM[1)]	Mina-~~DAT~~ hug-psv	0	0
	undetermined, no CM[1)]	Ciwu-~~NOM~~ hug-psv	0	0
Ditransitive	recipient-DAT only[2)]	Ciwu-DAT give	234	5
	SUM		6,902	143

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 the DAT is often used as an indicator of a recipient in the active and thus a potential competitor of the agent-DAT pairing in the passive.

ACC = accusative case marker; DAT = dative marker; NOM = nominative case marker

Methods

- Input composition
  - Artificial set of input based on the characteristics of caregiver input in CHILDES
  - Schematised input comprising two layers in a pair
    - **morpho-syntactic** layer specifying formal properties of the pattern
    - **semantic-functional** layer indicating thematic roles of arguments and functions of markers

(4) Example of input: canonical active transitive, no omission

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

- not abstract categories but heuristics—strategic and provisional knowledge which is acquired probabilistically through exposure—that a learner employs in the course of acquisition
  - a word with a marker stands for an entity
  - a word at the end of a sentence refers to an action

ACC = accusative case marker; NOM = nominative case marker

Methods

$$\text{Best Construction } (nCx) = \underset{eCx}{\text{argmax}}\ P(eCx \mid nCx)$$

- Model training
  - Adding a new input item to an existing group of constructions that had the most similar characteristics to the item
  - Laplace smoothing
    - the Laplace estimator added the value of 1 as the Laplace value to the original frequency value so that the probability of occurrence did not become zero and thus incalculable
  - Transitional probability: a series of conditional probabilities from the first item to the last item within a specific pattern; incremental

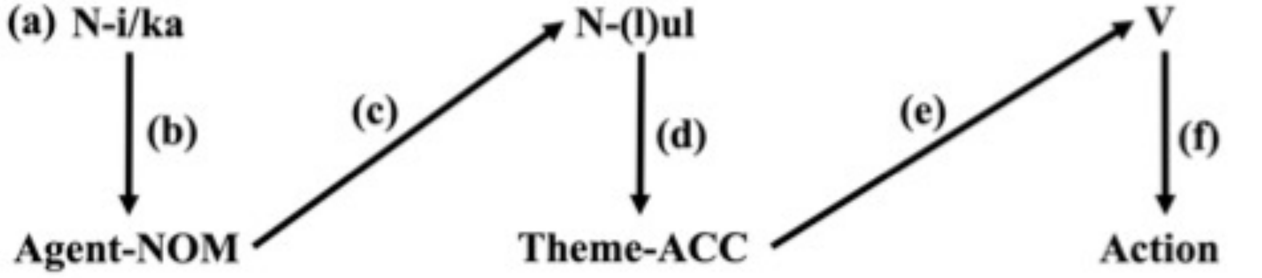

Figure 1. Schematic display of how to calculate transitional probability (canonical active transitive with no omission of arguments and case-marking)

ACC = accusative case marker;
NOM = nominative case marker

Methods

- To clarify,
  - Modelled a child learner after the age of one or two, following the age range of the children in CHILDES → employed frequency information in the caregiver input as the initial priors of our learning model, instead of creating a *tabula rasa* model from scratch

  - Intended to model the development of linguistic knowledge about clause-level constructions themselves → devising a set of schematised input instead of using concrete lexical words, for the model training
    - supported by an *early abstraction* account (i.e., the early emergence of abstract knowledge and yet still requiring considerable amount of exposure for the maturation of that knowledge)

Results & Discussion

- The only pattern whose degree of clustering was constantly increasing as learning proceeded

Table 2. By-construction posterior probability per learning

Type (example)	Posterior probability			
	1	3	5	10
Canonical active transitive, no omission	0.475	0.625	0.675	0.799
Scrambled active transitive, no omission	0.004	0.002	0.001	0.001
Canonical active transitive, no ACC	0.020	0.010	0.006	0.004
Canonical active transitive, no NOM	0.001	0.001	< 0.001	< 0.001
Active transitive, agent-NOM only	0.068	0.034	0.023	0.012
Active transitive, theme-ACC only	0.179	0.089	0.060	0.033
Active transitive, agent only, no case-marking	0.004	0.002	0.001	0.001
Active transitive, theme only, no case-marking	0.178	0.184	0.186	0.107
Active transitive, undetermined, no case-marking	0.002	0.001	< 0.001	< 0.001
Suffixal passive, theme-NOM only	0.029	0.015	0.010	0.005
Suffixal passive, agent-DAT only	0.001	0.001	< 0.001	< 0.001
Suffixal passive, theme only, no case-marking	0.002	0.001	< 0.001	< 0.001

Note. The other constructional patterns not listed in this table converged upon zero probability immediately after the 1st learning. The ditransitive pattern only with the recipient-DAT pairing does not fall into a transitive event, so we excluded the pattern in this table. For the sake of readers, it achieved the posterior probability of 0.035 and 0.036 after the 1st and the 10th of learning, respectively.

ACC = accusative case marker; DAT = dative marker; NOM = nominative case marker

Results & Discussion

- The posterior probability was not the highest
  - The most frequent pattern for a transitive event attested in the input
- The posterior probability increased until the 5th learning but decreased afterwards
  - The third most frequent pattern for a transitive event attested in the input
- Their growth was somehow inhibited, possibly by the growth of the clustering for the fully-equipped pattern

Table 2. By-construction posterior probability per learning

Type (example)	Posterior probability			
	1	3	5	10
Canonical active transitive, no omission	0.475	0.625	0.675	0.799
Scrambled active transitive, no omission	0.004	0.002	0.001	0.001
Canonical active transitive, no ACC	0.020	0.010	0.006	0.004
Canonical active transitive, no NOM	0.001	0.001	< 0.001	< 0.001
Active transitive, agent-NOM only	0.068	0.034	0.023	0.012
Active transitive, theme-ACC only	0.179	0.089	0.060	0.033
Active transitive, agent only, no case-marking	0.004	0.002	0.001	0.001
Active transitive, theme only, no case-marking	0.178	0.184	0.186	0.107
Active transitive, undetermined, no case-marking	0.002	0.001	< 0.001	< 0.001
Suffixal passive, theme-NOM only	0.029	0.015	0.010	0.005
Suffixal passive, agent-DAT only	0.001	0.001	< 0.001	< 0.001
Suffixal passive, theme only, no case-marking	0.002	0.001	< 0.001	< 0.001

Note. The other constructional patterns not listed in this table converged upon zero probability immediately after the 1st learning. The ditransitive pattern only with the recipient-DAT pairing does not fall into a transitive event, so we excluded the pattern in this table. For the sake of readers, it achieved the posterior probability of 0.035 and 0.036 after the 1st and the 10th of learning, respectively.

ACC = accusative case marker; DAT = dative marker; NOM = nominative case marker

Results & Discussion

- The passive patterns: cue competition involving case-marking and verbal morphology
  - The growth of this pattern may have been suppressed by its corresponding pattern—the active transitive with only the agent-NOM pairing
    - typical case-marking (the NOM indicating the agent)
    - typical verbal morphology (no active morphology)

Table 2. By-construction posterior probability per learning

Type (example)	Posterior probability			
	1	3	5	10
Canonical active transitive, no omission	0.475	0.625	0.675	0.799
Scrambled active transitive, no omission	0.004	0.002	0.001	0.001
Canonical active transitive, no ACC	0.020	0.010	0.006	0.004
Canonical active transitive, no NOM	0.001	0.001	< 0.001	< 0.001
Active transitive, agent-NOM only	0.068	0.034	0.023	0.012
Active transitive, theme-ACC only	0.179	0.089	0.060	0.033
Active transitive, agent only, no case-marking	0.004	0.002	0.001	0.001
Active transitive, theme only, no case-marking	0.178	0.184	0.186	0.107
Active transitive, undetermined, no case-marking	0.002	0.001	< 0.001	< 0.001
Suffixal passive, theme-NOM only	0.029	0.015	0.010	0.005
Suffixal passive, agent-DAT only	0.001	0.001	< 0.001	< 0.001
Suffixal passive, theme only, no case-marking	0.002	0.001	< 0.001	< 0.001

Note. The other constructional patterns not listed in this table converged upon zero probability immediately after the 1st learning. The ditransitive pattern only with the recipient-DAT pairing does not fall into a transitive event, so we excluded the pattern in this table. For the sake of readers, it achieved the posterior probability of 0.035 and 0.036 after the 1st and the 10th of learning, respectively.

ACC = accusative case marker; DAT = dative marker; NOM = nominative case marker

Results & Discussion

- The passive patterns: cue competition involving case-marking and verbal morphology
  - the growth of this pattern may have been suppressed by the ditransitive with only the recipient-DAT pairing
    - case-marking (the DAT indicating the recipient is more frequent than the DAT indicating the agent)
    - verbal morphology (verb with no morphology is more frequent than verb with passive suffixes)

Table 2. By-construction posterior probability per learning

Type (example)	Posterior probability			
	1	3	5	10
Canonical active transitive, no omission	0.475	0.625	0.675	0.799
Scrambled active transitive, no omission	0.004	0.002	0.001	0.001
Canonical active transitive, no ACC	0.020	0.010	0.006	0.004
Canonical active transitive, no NOM	0.001	0.001	< 0.001	< 0.001
Active transitive, agent-NOM only	0.068	0.034	0.023	0.012
Active transitive, theme-ACC only	0.179	0.089	0.060	0.033
Active transitive, agent only, no case-marking	0.004	0.002	0.001	0.001
Active transitive, theme only, no case-marking	0.178	0.184	0.186	0.107
Active transitive, undetermined, no case-marking	0.002	0.001	< 0.001	< 0.001
Suffixal passive, theme-NOM only	0.029	0.015	0.010	0.005
Suffixal passive, agent-DAT only	0.001	0.001	< 0.001	< 0.001
Suffixal passive, theme only, no case-marking	0.002	0.001	< 0.001	< 0.001

Note. The other constructional patterns not listed in this table converged upon zero probability immediately after the 1st learning. The ditransitive pattern only with the recipient-DAT pairing does not fall into a transitive event, so we excluded the pattern in this table. For the sake of readers, it achieved the posterior probability of 0.035 and 0.036 after the 1st and the 10th of learning, respectively.

ACC = accusative case marker; DAT = dative marker; NOM = nominative case marker

Results & Discussion

- Inconsistency
  - Considering the overall number of the constructional patterns that the children produced (143 instances), they seemed to prefer the three patterns, all of which engage in 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 3. Inconsistency between corpus analysis and the simulation

Type	Frequency of occurrence (corpus analysis)		Posterior probability (simulation; 10th learning)
	Caregiver input (#)	Child production (#)	
Active transitive, agent-NOM only	935	21	0.012
Canonical active transitive, no ACC	268	14	0.004
Suffixal passive, theme-NOM only	407	9	0.005

ACC = accusative case marker; NOM = nominative case marker

Results & Discussion

- Inconsistency
  - Why?
    - our computational model faithfully followed the construction-based distributional properties attested in the caregiver input
      - Considering that our learning model proceeded with transitional probabilities that care about both construction-wise characteristics and case-marking facts, it is reasonable to think that the model well-responded to the construction frequency as much as the form-function mapping of case-marking in the input

Table 3. Inconsistency between corpus analysis and the simulation

Type	Frequency of occurrence (corpus analysis)		Posterior probability (simulation; 10^{th} learning)
	Caregiver input (#)	Child production (#)	
Active transitive, agent-NOM only	935	21	0.012
Canonical active transitive, no ACC	268	14	0.004
Suffixal passive, theme-NOM only	407	9	0.005

ACC = accusative case marker; NOM = nominative case marker

Results & Discussion

- Inconsistency
  - Why?
    - The children, however, may have relied more on the reliable and available form–function mapping of NOM for a transitive event observed in caregiver input
      - NOM ~ agent; NOM occur more frequently in the initial position than in the non-initial position
      - These characteristics allow high cue validity for this particular mapping (Bates & MacWhinney, 1987), leading the children to primarily (and strongly) deploy NOM to indicate the actor of a transitive event

Table 3. Inconsistency between corpus analysis and the simulation

Type	Frequency of occurrence (corpus analysis)		Posterior probability (simulation; 10th learning)
	Caregiver input (#)	Child production (#)	
Active transitive, agent-NOM only	935	21	0.012
Canonical active transitive, no ACC	268	14	0.004
Suffixal passive, theme-NOM only	407	9	0.005

ACC = accusative case marker; NOM = nominative case marker

Results & Discussion

- Inconsistency
  - What about the suffixal passive with only the theme–NOM pairing?
    - Influence of lexical items:
      - the way that the child produced this particular constructional pattern was tied to several verbs
      - Child production may have been limited to less-abstract and narrow-range schemata in its initial phase (verb-centred frames)

Table 3. Inconsistency between corpus analysis and the simulation

Type	Frequency of occurrence (corpus analysis)		Posterior probability (simulation; 10th learning)
	Caregiver input (#)	Child production (#)	
Active transitive, agent-NOM only	935	21	0.012
Canonical active transitive, no ACC	268	14	0.004
Suffixal passive, theme-NOM only	407	9	0.005

ACC = accusative case marker; NOM = nominative case marker

Results & Discussion

- To recap, there are two particular motivations for our simulation
  - We modelled a child learner after the age of one or two, following the age range of the children in CHILDES
    - employed frequency information in the caregiver input as the initial priors of the learning model
  - We modelled the development of linguistic knowledge about clause-level constructions in their entirety
    - devised the schematised input with a pair of two abstract layers, instead of using content words attested in the caregiver input

Results & Discussion

- Limitations
  - The fact that we composed input without lexical information renders it impossible for the model to capture this lexically-tied factor to the extent that human learners do when acquiring constructional knowledge
  - 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

Results & Discussion

- Nevertheless, our findings show…
  - some nice compatibility of model performance with the child production
    - The distributional properties of the constructional patterns for a transitive event and the particular characteristics of form–function associations involving case-marking and verbal morphology dedicated to these constructions found in the caregiver input yielded the model performance that was largely consistent with the child production, despite no individual support from lexical information
    - provides a nice approximation to how constructional knowledge grows and changes in Korean-speaking children's conceptual space as a response of (i) construction frequency within the amount of input attested and (ii) form–function correlations involving the core structural properties consisting of the target construction types (cf. Competition Model)

Results & Discussion

- Our findings also highlight …
  - the status of abstract form-function mapping, as a form of construction which is independent of individual lexical items, as a psychological reality in language development
    - we exclusively considered information about constructions themselves, with particular emphasis on properties of construction-wise distribution and structural components characterising each construction type.
    - our novel approach affords us to effectively examine the extent to which children cope with knowledge about clause-level constructions in their entirety during learning

References (selected)

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.

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

Alishahi, A., & Stevenson, S. (2008). *A computational model of early argument structure acquisition. Cognitive Science, 32*(5), 789–834.

Bannard, C., Lieven, E., & Tomasello, M. (2009). Modeling children's early grammatical knowledge. *Proceedings of the National Academy of Sciences, 106*(41), 17284–17289.

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. (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.

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

Lieven, E. (2010). Input and first language acquisition: Evaluating the role of frequency. *Lingua, 120*(11), 2546–2556

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

Perfors, A., Tenenbaum, J. B., Griffiths, T. L., & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition, 120(*3), 302–321.

References (selected)

Perfors, A., Tenenbaum, J. B., & Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition, 118(*3), 306–338.

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.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926–1928.

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

Thank you for your listening!