Bayesian modelling of clause-level constructional knowledge for Korean-speaking preschool children

Gyu-Ho Shin (University of Hawai'i at Mānoa) & Seongmin Mun (Université Paris Ouest Nanterre La Défense)

BACKGROUND

Usage-based constructionist approaches

- Language development as interactions between frequency and domain-general learning capacities (e.g., Goldberg, 2019; Tomasello, 2003)
- *Q: how do we appropriately represent developmental trajectories involving clusters of form-function pairings (i.e., constructions)?*

Bayesian-inference-based simulation

- Assumption: human learning involves one's updated beliefs based on previous experience
- Studies focused mostly on English (e.g., Alishahi & Stevenson, 2008; Barak et al., 2016; Perfors et al., 2011)
- Q: *to what extent are the implications of computational simulations generalisable across languages?*

Active transitives & suffixal passives in Korean

- Korean: SOV language with overt case-marking
- Clause-level constructions expressing a transitive event

Active transitive			
Canonical	N-NOM	N-ACC	V
Scrambled	N-ACC	N-NOM	V

Suffixal passive			
Canonical	N-NOM	N-DAT	V-PSV
Scrambled	N-DAT	N-NOM	V-PSV

- Language-specific properties
  - Arguments / case markers can be omitted if they are inferable from the context (Sohn, 1999)

Argument + case-marking omission

~~Ciwu-ka~~	Minho-lul	cap-ass-ta.
~~Ciwu-NOM~~	Minho-ACC	catch-PST-SE

'Ciwu caught Minho.'

Case marking omission

Ciwu-ka	Minho-~~lul~~	cap-ass-ta.
Ciwu-NOM	Minho-~~ACC~~	catch-PST-SE

'Ciwu caught Minho.'

- Form-function pairings involving case-marking
  - Asymmetric degree of association between form and function

- Passive morphology
  - Rarely attested in input; morphologically irregular; overlap in morphological causatives (Shin, 2020; Sohn, 1999; Yeon, 2015)

RQ *Given language-specific properties in Korean, how a Bayesian learner formulates knowledge about active transitives and suffixal passives?*

Abbreviation: ACC = accusative case marker; DAT = dative marker; N = noun; NOM = nominative case marker; PST = past tense marker; SE = sentence ender; V = verb

BAYESIAN SIMULATION

Input composition

- All constructional patterns expressing a transitive event found in caregiver input in CHILDES (MacWhinney, 2000)

Type	Example	Frequency (#)
Canonical active transitive	police-NOM thief-ACC catch	1,757
Scrambled active transitive	thief-ACC police-NOM catch	51
Canonical suffixal passive	thief-NOM police-DAT catch-psv	2
Scrambled suffixal passive	police-DAT thief-NOM catch-psv	1
Canonical active transitive, no ACC	police-NOM thief-~~ACC~~ catch	268
Canonical active transitive, no NOM	police-~~NOM~~ thief-ACC catch	19
Scrambled active transitive, no ACC	thief-~~ACC~~ police-NOM catch	6
Scrambled active transitive, no NOM	thief-ACC police-~~NOM~~ catch	0
Canonical suffixal passive, no DAT	thief-NOM police-~~DAT~~ catch-psv	0
Canonical suffixal passive, no NOM	thief-~~NOM~~ police-DAT catch-psv	0
Scrambled suffixal passive, no DAT	police-~~DAT~~ thief-NOM catch-psv	0
Scrambled suffixal passive, no NOM	police-DAT thief-~~NOM~~ catch-psv	0
Active transitive, actor-NOM only	police-NOM catch	935
Active transitive, undergoer-ACC only	thief-ACC catch	1,938
Ditransitive, recipient-DAT only	Lee-DAT send	234
Suffixal passive, undergoer-NOM only	thief-NOM catch-psv	407
Suffixal passive, actor-DAT only	police-DAT catch-psv	13
SUM		5,631

- Schematised pairings of morpho-syntactic and semantic-functional properties; indexing for canonicity

Example of input: canonical active transitive

Morpho-syntactic layer	N_1-i/ka_1	N_2-(l)ul_2	V_3
Semantic-functional layer	Actor_1-NOM_1	Undergoer_2-ACC_2	Action_3

※ N and V represent (probabilistically acquired) heuristics of noun and verb, respectively

Model training

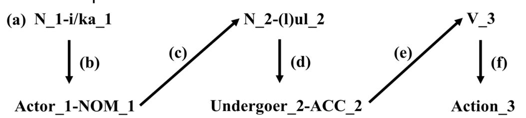
- Frequency of constructional patterns in caregiver input → initial priors for learning
- Learning algorithm (adapted from Alishahi & Stevenson, 2008)
  - A new input *nCx* is classified as an existing construction *eCx*, ranging over the indices of all the constructions in the model, with the maximum probability given *nCx*

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

  - Posterior probability is *proportional* to multiplication of conditional probabilities associated with *eCx* and the prior of *eCx*

$$P(eCx \mid nCx) \propto P(nCx \mid eCx) * P(eCx)$$

  - Laplace smoothing to prevent the probability from converging upon zero
- Two types of probability information
  - Constructional probability: probabilities of individual patterns
  - Transitional probability: conditional probabilities of constructional components within each pattern



Model performance

- Posterior probabilities of constructional patterns at every learning phase (one to 30)
  - as a proxy for the degree of clustering for these constructions

RESULTS & DISCUSSION

By-pattern posterior probabilities

- *Dominance* of several patterns over the others

Type	Caregiver input (#)	Posterior probability per learning		
		1	5	30
Canonical active transitive	1,757	0.454	0.550	0.588
Scrambled active transitive	51	0.005	0.002	< 0.001
Canonical suffixal passive	2	< 0.001	< 0.001	< 0.001
Scrambled suffixal passive	1	< 0.001	< 0.001	< 0.001

※ mirrored distributional nature of child production (cf. Shin, 2020)

→ *Inhibitory effects* on the growth of the related patterns

Type)	Caregiver input (#)	Posterior probability per learning		
		1	5	30
Canonical active transitive, no ACC	268	0.024	0.008	0.002
Canonical active transitive, no NOM	19	0.002	0.001	< 0.001
Active transitive, actor-NOM only	935	0.083	0.028	0.005
Active transitive, undergoer-ACC only	1,938	0.351	0.355	0.357
Suffixal passive, undergoer-NOM only	407	0.036	0.012	0.002
Suffixal passive, actor-DAT only	13	0.001	< 0.001	< 0.001

※ The other patterns converged upon zero probability immediately after the 1st learning

- *Inconsistency* between simulation and child production

Type	Caregiver input (#)	Child production (#)	Posterior probability (30th)
Active transitive, actor-NOM only	935	21	0.005
Canonical active transitive, no ACC	268	14	0.002
Suffixal passive, undergoer-NOM only	407	9	0.002

- NOM-related patterns
- Possible reasons
  - Influences of case-marking (i.e., NOM is used exclusively as an indicator of the actor in transitive patterns; cf. Shin, 2020)
  - Non-transitive partial utterances (with various noun-marker combinations) not considered in the current simulation
  - Lexical items tied to specific constructional patterns in children's utterances

Together, our findings…

- ✓ ***support the idea that clause-level constructional knowledge grows through an interplay between input properties and domain-general learning capacities***
- ✓ ***adds to cross-linguistic evidence for the effectiveness of Bayesian modelling on representing human learning***

REFERENCES

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

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.

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

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

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

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

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

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

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

Presented at the IMPRS Conference 2020 (03-05 JUN 2020)

Corresponding Author: Gyu-Ho Shin (*gyuhoshin@gmail.com*)