

Korean-speaking children’s constructional knowledge about transitive events: Corpus analysis and Bayesian modelling

This study investigates Korean-speaking children’s knowledge about argument structure constructions involving a transitive event—active transitives and suffixal passives—through corpus analysis and Bayesian modelling. Usage-based constructionist approaches assume that language emerges and grows via interactions of frequency and domain-general learning capacities (Goldberg, 2019; Tomasello, 2003). The issue is how to appropriately represent developmental trajectories involving clusters of form-function pairings (constructions; Goldberg, 1995) based on exposure, together with non-linguistic forces from cognitive-psychological factors. Korean, which is typologically different from major languages under investigation regarding this issue, provides an intriguing testbed due to language-specific properties such as agglutination, scrambling, and omission of sentential components.

A canonical active transitive (1a) occurs with the nominative-marked agent, followed by the accusative-marked theme, with no active verbal morphology *per se*. A canonical suffixal passive (2a) occurs with the nominative-marked theme, followed by the dative-marked agent, with passive morphology attached to a verb. These patterns can be scrambled (1b; 2b).

Oftentimes, omission applies to a marker or an argument and a marker altogether.

(1a)	agent-nominative	theme-accusative	V
(1b)	theme-accusative	agent-nominative	V
(2a)	theme-nominative	agent-dative	V-passive
(2b)	agent-dative	theme-nominative	V-passive

Study 1: Corpus analysis. We analysed caregiver input and child production attested in CHILDES (MacWhinney, 2000), the largest and open-access child corpora in Korean comprising nine caregivers and four children whose ages range from 1;3 to 3;10. Automatic analysis of this database revealed that (1) the rates of constructional patterns produced by the children mirrored those uttered by the caregivers to a considerable degree, and (2) the caregivers’ use of case-marking was skewed towards single form-function mapping (despite the multiple form-function associations that the markers manifest in actual language use).

Study 2: Bayesian modelling. Based on the characteristics of the caregiver input in CHILDES, we modelled our Bayesian learner, by adapting the algorithms of Alishahi and Stevenson (2008). The model learnt instances of the two construction types as schematised input—pairings of morpho-syntactic and semantic-functional properties involving these constructions (with varying degrees of omission of case-marking and/or arguments)—based on the distributional properties of CHILDES. Posterior probabilities of the patterns per learning phase were measured to estimate the degree of clustering for these constructions. We found dominance of several patterns, also occupying a large portion of the caregiver input, and their inhibitory effects on the growth of their related patterns. This largely mirrored the distributional properties of child production found in CHILDES. Some inconsistencies between the child production and the simulation were related to the patterns involving the nominative case marker, which is attributable to (i) the exclusive status of this marker in a transitive event (i.e., a very reliable cue for the agent and vice versa) and (ii) the impact of partial utterances, together with concrete lexical items, not considered in the simulation.

Together, our findings extend the current understanding of how Korean-speaking children shape clause-level constructional knowledge as a function of input properties and domain-general learning capacities.

Keywords: construction, corpus analysis, Bayesian modelling