

Speakers' preference for more versus less-transparent causatives: Computational modeling, grammaticality judgment and production data from English, Hebrew, Hindi, Japanese, K'iche' Mayan and Balinese.

Ben Ambridge
University of Liverpool, UK
ESRC International Centre for Language and Communicative Development (LuCiD)

Laura Doherty
University of Liverpool, UK

Ramya Maitreyee
University of Liverpool, UK

Tomoko Tatsumi
Kobe University, Japan

Shira Zicherman,
Hebrew University of Jerusalem, Israel

Pedro Mateo Pedro
Universidad del Valle de Guatemala

Ayuno Kawakami
University of Liverpool, UK

Amy Bidgood
University of Salford, UK

Clifton Pye
University of Kansas

Bhuvana Narasimhan
University of Colorado, Boulder

Shira Zicherman
Inbal Arnon
Dani Bekman
Amir Efrati
Hebrew University of Jerusalem

Sindy Fabiola Can Pixabaj
Mario Marroquín PelízMargarita Julajuj Mendoza,
Universidad del Valle de Guatemala

Soumitra Samanta
University of Liverpool, UK
ESRC International Centre for Language and Communicative Development (LuCiD)

Colin Bannard
University of Liverpool, UK

Seth Campbell
University of Calgary

Stewart McCauley
University of Iowa

Ruth Berman
Tel Aviv University

Dipti Misra Sharma
Indian Institute of Information Technology, Hyderabad

Rukmini Bhaya Nair
Indian Institute of Technology, Delhi

Kumiko Fukumura
University of Strathclyde

This work investigates the ways in which six different natural languages use morphosyntactic marking to mark difference degrees of agency, and the continuum from intentional and/or directly-caused to accidental and/or indirectly-caused events. Many of the world's languages (Shibatani & Pardeshi, 2002, discuss 38 examples) have (at least) two causative structures, which particular verbs prefer to a greater or lesser degree: (1) a *more-transparent* structure with the verb *cause/make/do* or a morpheme that is often a historically grammaticalized form of that verb (e.g., Japanese *-(s)ase*), and (2) a *less-transparent* structure that marks causation more idiosyncratically (e.g., by using a form that is either indistinguishable from a non-causative or stem form, or similar to such a form, but with a vowel or consonant change that is only partially predictable):

- (1) Someone made the truck break [More transparently-marked causation]
- (2) Someone broke the truck [Less transparently-marked causation]

The formal realizations of these different forms vary from language to language. English, as illustrated by the examples above, relies primarily on syntax. Japanese, Hindi and K'iche rely primarily on morphology, in the form of a more-transparent causative marker (*-(s)ase*, *-aa*, and *-(i)sa-j* respectively) and various types of less-transparent stem-change. For Hebrew, the root is defined as a three-consonant (C.C.C) pattern (e.g., *sh.b/v.r* for BREAK), which forms a verb only when it is inserted into a binyan template; in this case either the dedicated causal binyan *hiCCiC* (e.g., *hishbir*) or the appropriate general transitive binyan: CaCaC (e.g., *shavar*) CiCeC or hiCaCeC, for the more- and less-transparent forms respectively.

The aim of this study was to investigate how speakers learn which verbs prefer which causative form (more-/less-transparent), and to what extent. First, 20 adult native speakers of each language rated each of 60 actions for four semantic properties relative to the notion of agency (from Shibatani & Pardeshi, 2002):

Event-merge: The extent to which the causing and caused event are two separate events or merge into a single event that happens at a single time and a single point in space

Autonomy of the causee

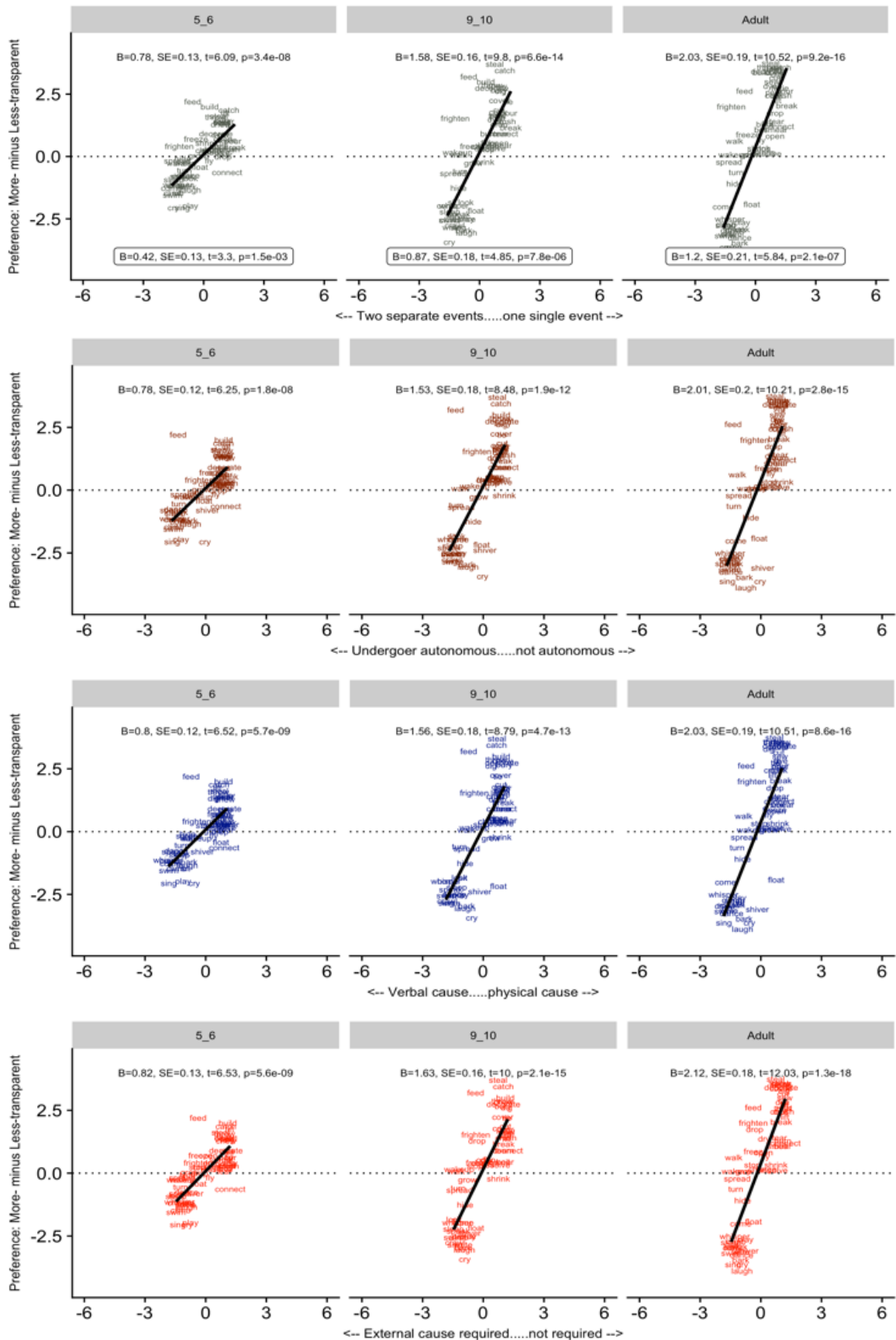
Requires: Whether the caused event requires a causer

Directive: Whether causation is directive (e.g., giving an order) or physical

These ratings, together with corpus frequency information, were to build a computational model of speakers' causative preferences for each language. For every language except K'iche', this model significantly predicted speakers' preference for the more- versus less-direct causative form across 60 verbs, as assessed using continuous grammaticality judgments (adults, children aged 5-6 and 9-10), binary grammaticality judgments (4-5) and elicited production (4-5 and 5-6) ($N=48$ participants per age group per study). The model was also able to generalize to unseen verbs on the basis of their semantics (when a randomly-selected half of the verbs were presented to the model for the first time at test), and even to ratings from an entirely new language, on which the model had not been trained (Balinese). We conclude that the morphosyntactic realization of causativity is determined primarily by the degree of conceptual merging between the causing and caused event (Shibatani & Pardeshi, 2002), and that children have some awareness of this fact by at least 4-5 years of age.

Shibatani, M. & Pardeshi, P. (2002). The causative continuum. In *The Grammar of Causation and Interpersonal Manipulation*, Masayoshi Shibatani (ed.), pp. 85–126. Amsterdam: John Benjamins.

Figure 1. Example data from English. Participants' preference for the more-transparent (periphrastic *make* causative) over the less-transparent (transitive causative) form for each verb, as a function of the semantic predictors of (a) Event-Merge, (b) Autonomy of the causee, (c) Whether causation is directive or physical and (d) whether the caused event requires a causer.



Note: The values shown for each predictor are from nonpartial (single predictor) models only