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 Pelíz, Margarita Julajuj Mendoza, Universidad del Valle de Guatemala; Soumitra Samanta, University of Liverpool, UK, 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 project has received funding from the European Research Council (ERC) under the European Union's research and innovation programme (grant agreement no 681296: CLASS

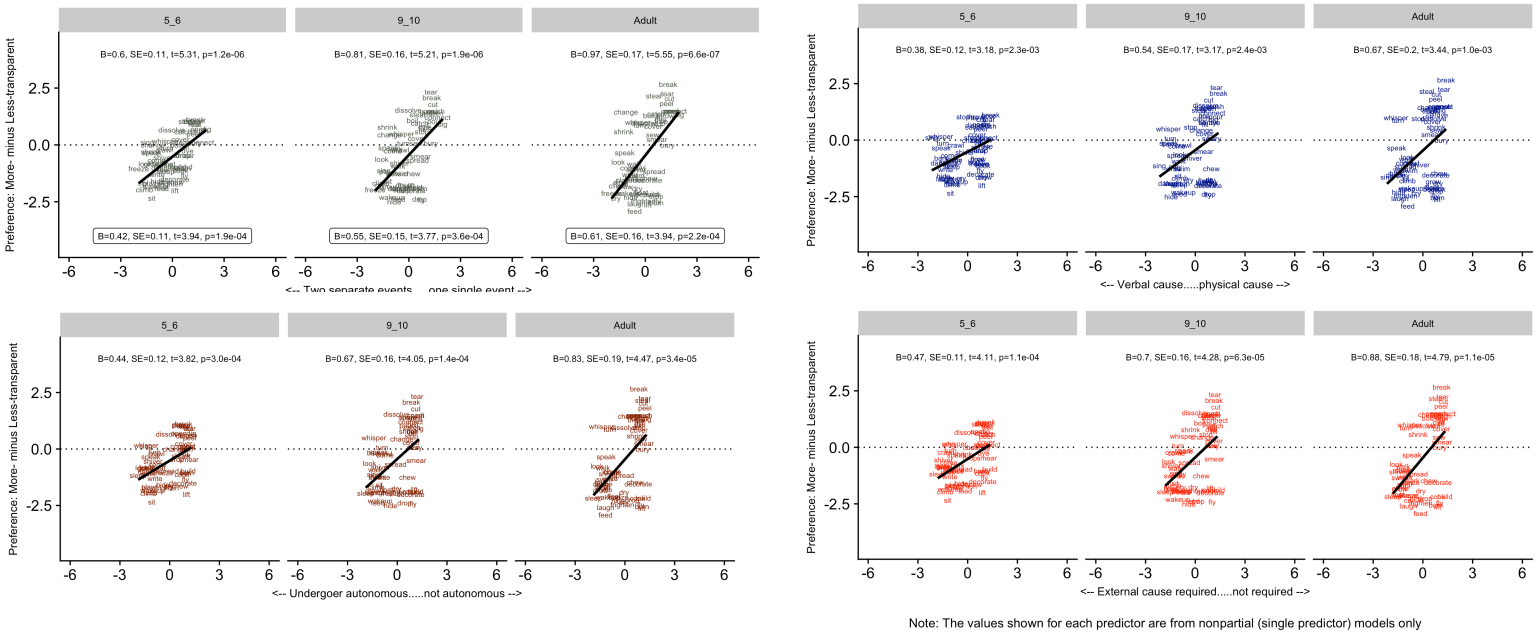
Many languages have more- and less-transparently marked causatives that indicate, respectively, indirect and direct causation (Hall, 1965; Bar-Asher Siegal, Bassel & Hagmayer, 2021).

Language Root: break	MORE TRANSPARENT "Morphological", regular, predictable, indirect	LESS TRANSPARENT "Lexical", irregular semi-(un)predictable, direct
K'iche' paax	-isa-j *Le achi x-0-u-**paax-isa-j** le qumub'al	zero marking Le achi x-0-u-**pax-ii-j** le qumub'al
Japanese koware	-(s)ase *Dareka ga torakku o **koware-sase-ru**	"lexical" Dareka ga torakku o **kowas-u**
Hindi TUT	-aa *kisi-ne Trak-ko **TUT-aa**	Vowel change kisi-ne Trak-ko **toD**
English break	Periphrastic ?Someone **made** the truck **break**	Transitive Someone broke the truck
Hebrew sh.b/v.r	Hif'il binyan *Mishehu **hishbir** et ha-masa'it	Pa'al / pi'el mishehu **shavar** et ha-masa'it

But what exactly counts as (in)direct causation, and how well does this predict speakers' preference the less(/more) transparent causative?

For each language, 20 adults rated each of 60 causal animations (with the causer hidden) for each of four semantic properties. **(a) Event Merge:** (b) **Autonomy** of the causee; (c) Extent caused event **Requires** a causer, and (d) Extent causation is **Directive** vs physical.

Ratings used (controlling for frequency) to predict speakers' preference for the more- versus less-direct causative form across 60 corresponding verbs (N=48 per age group per study) in **(a) Continuous grammaticality judgments (adults, children aged 5-6 and 9-10),** (b) 2. Binary grammaticality judgments (4-5), (c) Elicited Production (4-5 and 5-6). Hindi only shown:



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

Discriminative learning models trained to map from verb→causative type on the basis of (a) corpus co-currence frequencies and (b) these semantic ratings achieve significant correlations with (a) these continuous judgments (*r*=0.75 in most cases) https://www.sciencedirect.com/science/article/pii/S0010027720301293#s9915 (b) binary grammaticality judgment data from children aged 4;0-5;0 (r=0.5-0.6), and (c) elicited-production data from children aged 4;0-5;0 and 5;6-6;6 (*r*=0.75-0.85): https://osf.io/7v8m5/ (submitted to Open Research Europe).