LINGUISTICS

From language development to language evolution: A unified view of human lexical creativity

Thomas Brochhagen[1]*, Gemma Boleda[1,2], Eleonora Gualdoni[1], Yang Xu[3]

A defining property of human language is the creative use of words to express multiple meanings through word meaning extension. Such lexical creativity is manifested at different timescales, ranging from language development in children to the evolution of word meanings over history. We explored whether different manifestations of lexical creativity build on a common foundation. Using computational models, we show that a parsimonious set of semantic knowledge types characterize developmental data as well as evolutionary products of meaning extension spanning over 1400 languages. Models for evolutionary data account very well for developmental data, and vice versa. These findings suggest a unified foundation for human lexical creativity underlying both the fleeting products of individual ontogeny and the evolutionary products of phylogeny across languages.

Humans often need to talk about new entities and concepts but must rely on a limited vocabulary of known words to do so. A common solution to overcoming this bottleneck is the creative use of single words to express multiple meanings through a process known as word meaning extension (*1*–*3*).

In linguistic development, Vygotsky, among others, documented word meaning extension in young children, noting that a word such as "quah" can be overextended to express "a duck," "water," "liquid," or "a coin with an eagle on it" (*4*–*6*). At the individual level, child overextension is transient: It occurs during the early stages of life and vanishes in later language development (*7*) (the definition of overextension and other key terms can be found in Table 1). By contrast, at the population level in language evolution, more stable forms of lexical creativity become entrenched in language after longer periods of time because of cultural transmission. Colexification—the phenomenon by which related meanings (such as "finger" and "toe") are expressed with the same word (e.g., "dit" in Catalan)—can be a product of this process and is attested across languages (*8*–*11*). Similarly, words can also acquire new meanings over time through semantic change (*1*, *12*–*14*). For example, the word "mouse" was extended to refer to a computer device because of the visual similarity between the device and a rodent.

Despite these differences in the manifestations of lexical creativity across levels (individual versus population) and timescales (short versus long), we posit that children and language users in general tackle the same fundamental task: to extend known words to referents that lack a word by relating those referents to the known words' current meanings. If this is true, we expect the kinds of overextensions that appear in child development to be similar to the kinds of meaning extensions attested in language evolution because both build on a common foundation grounded in human experience and cognition. We developed a computational framework to test this idea and have shown that it receives support in a large-scale analysis.

Developmental and evolutionary phenomena pertaining to lexical creativity have been studied by different research communities. Research in developmental psychology suggests that child overextension relies on the ability to identify similarity relations among concepts. This ability has been shown to draw on multiple types of semantic knowledge by using perceptual, action-functional, affective, and contextual information (*6*, *15*). Similarly, recent work has demonstrated that visual, taxonomic, and associative information can jointly explain a variety of child overextension patterns (*15*). Work in linguistics and cognitive psychology has similarly suggested that semantic relatedness plays a role in colexification (*10*, *11*, *16*–*18*) and historical semantic change (*1*, *3*, *13*, *19*, *20*). In this work, we integrated these separate lines of research by asking whether different forms of lexical creativity, from that of childhood to language evolution, rely on a common foundation.

Our study is relevant to the long-standing issue concerning the relation between linguistic development and the evolution of language. One locus of this issue is the possibility that ontogeny recapitulates phylogeny in language (*21*, *22*); that is, whether there are recurring patterns in child development that inform or reflect patterns in language evolution or vice versa. Previous work has suggested that ontogeny is shaped by a particular language, emphasizing learning at the individual level, whereas phylogeny is a product of lasting innovations, emphasizing language use at the population level (*23*). By contrast, recent studies have shown that cross-linguistic patterns can recur during individual language learning (*24*–*26*), that cultural evolution of linguistic structures can be recapitulated in the laboratory (*27*), and that some development data predict lexical evolution rates (*28*).

We investigated the relation between linguistic development at the individual level and language evolution at the population level by taking a different perspective informed by word

Table 1. Key terms and their definitions as operationalized in this study.

Term	Definition
Affectiveness	Measure of how pleasant, intense, and dominant a term is perceived to be (e.g., "sunshine" scores high on all three dimensions).
Associativity	Measure of how relatable two terms are (e.g., "key" and "door" are more closely associated than either is to "dog").
Colexification	Meanings expressed by the same word are colexified (e.g., Catalan uses a single word, "dit," to express both "finger" and "toe").
Lexical/lexicon	Relating to words and their meanings, or the knowledge thereof.
Ontogeny	An individual's developmental history (e.g., their language use from early acquisition to adulthood).
Overextension	Extended use of a known term to referents outside its normal category (e.g., a child saying "apple" for any round object or "dog" for any animal).
Phylogeny	Evolutionary history of a group (e.g., historical development of languages).
Semantics	Meaning, or the study thereof.
Semantic change	Historical change in word meaning (e.g., "mouse" taking on the new meaning "computer device").
Taxonomy	Hierarchical or tree-based classification (e.g., an apple is a fruit)
Visual similarity	Measure of resemblance based on visual features (e.g., how similar apples and balls look across many images).

[1]Department of Translation and Language Sciences, Universitat Pompeu Fabra, Barcelona, Spain. [2]Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain. [3]Department of Computer Science, Cognitive Science Program, University of Toronto, Toronto, Canada.
*Corresponding author. Email: thomas.brochhagen@upf.edu

meaning extension. There is only a small subset of meaning extensions that directly overlap between child development and products of language evolution (supplementary materials) (Fig. 1A). This might be taken as weak evidence for the idea that ontogeny at least partially recapitulates phylogeny or vice versa. Our approach aims at understanding the nonoverlapping, broader space of creative meaning extensions that is illustrated in Fig. 1A: not only the intersection, but the union of these phenomena.

We present a unified view of lexical creativity by hypothesizing that there is a latent common foundation that children and language users in general both build on when using words creatively. This common foundation relies on two components: first, grounded knowledge about objects, events, properties, and relations, such as objects having certain shapes or belonging to certain categories; and second, the use of this knowledge to link referents lacking a word with current meanings of known words based on similarities between the two. We thus tested the proposal that both the types of knowledge and the use of similarity in word meaning extension are shared in child overextension and products of lexical creativity from language users in general.

Figure 1B summarizes the computational framework we developed to test our proposal, which involves (i) explicitly defining a set of semantic knowledge types as proxies for the hypothesized common foundation and (ii) using them to make cross-predictions about products of lexical creativity in development and evolution. If the different forms of lexical creativity draw on different knowledge types or do so to very different degrees, we would expect minimal carryover between the phenomena. If instead our unified view is warranted, we expect good cross-predictability; that is, we expect that models built from child data will successfully account for data that are the product of language evolution and vice versa.

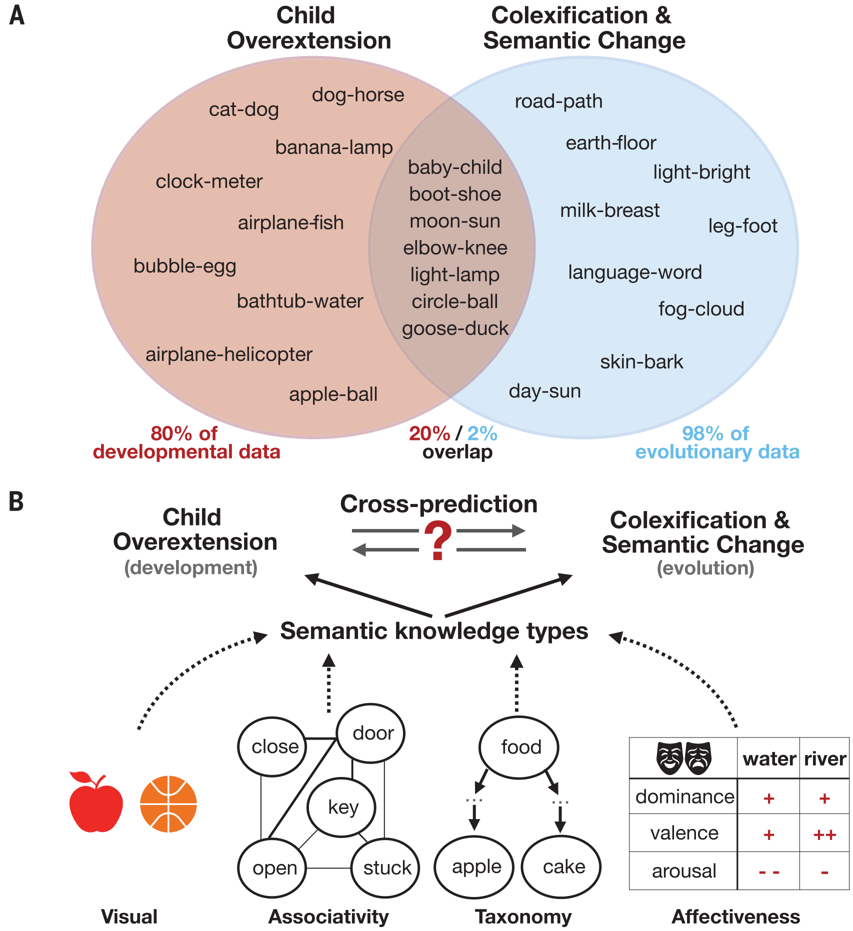

Fig. 1. Illustrations of the phenomenon of lexical creativity and the overall framework. (**A**) Examples of word meaning extension as a common form of lexical creativity. Each is an attested pair of meanings that are coexpressed by a word form. The circle's intersection shows examples of child overextensions (in development) that are also attested in lexicons of the world's languages (through evolution). Cases outside of this area are only attested at one timescale. (**B**) Framework for investigating the possibility of a common foundation in lexical creativity. Four semantic knowledge types are considered: visual similarity, associativity, taxonomy, and affectiveness. The framework enables cross-prediction between developmental and evolutionary phenomena.

Framework

We developed a framework that incorporates four semantic knowledge types discussed commonly in the literature in connection to child overextension, colexification, and semantic change (*6*, *10*, *11*, *15*, *18*, *29*, *30*): associativity, and similarities based on visual, taxonomic, and affective information (Fig. 1B). These are largely complementary in the information they provide (tables S1 to S3), but they are not exhaustive. We operationalized these knowledge types on the basis of English resources, a limitation to which we returned later. Materials and methods are available as supplementary materials.

We operationalized visual similarity using computationally derived visual representations of meanings. We followed a two-step procedure, drawing on existing work (*15*, *31*). First, we used a computer vision model (*32*) to produce representations for images of instances of meanings (e.g., for images of dogs for the meaning "dog"). Second, we averaged these instance representations, yielding average visual representations that we took as surrogate meaning representations. We used these average representations to calculate the visual similarity of different meanings.

We defined associativity in terms of how closely meanings are relatable in semantic memory. We quantified this using large-scale experimental data (*29*) that records the responses produced by subjects when prompted with a cue word (e.g., "dog" may elicit "cat," "bone," or "cuddly"). To obtain a measure of associativity, we transformed cue-response counts using the best method identified in the literature so far (*11*, *29*).

We took taxonomic similarity as a proxy for the categorical relatedness of meanings (e.g., "dog" and "cat" are taxonomically closer than either is to "key" or "love"). Following previous work on child overextension (*15*), we used a measure based on a large lexical database (*33*). The measure yields a score for the similarity of two meanings based on their closest common ancestor in a taxonomy.

We operationalized affectiveness as the similarity of affective experiential features such as emotional valence. More precisely, following (*30*), we quantified affective similarity between meanings as the cosine similarity of their vectors of ratings, built from two large-scale databases of affectiveness norms (*34*, *35*). These norms encompass ratings for valence, arousal, and dominance.

We analyzed three independent datasets that represent three phenomena of lexical

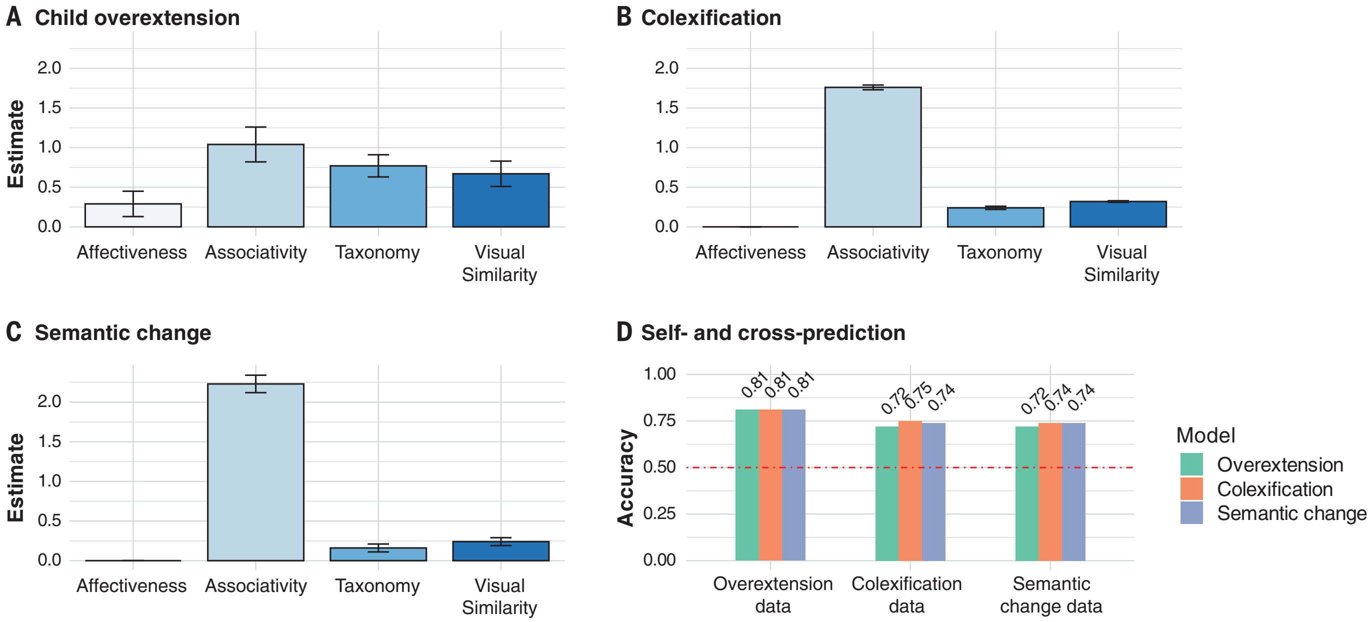

Fig. 2. Summary of main results. (**A** to **C**) Standardized estimates of the effect of knowledge types from best models of child overextension, colexification, and semantic change, respectively. (**D**) Accuracy of models when predicting new data (i.e., the fraction of correct predictions they make). Self-predictions (e.g., the colexification model's performance on colexification data) provide an upper bound for cross-predictions. The random baseline of 0.5 (dashed line) provides a lower bound. Ceiling and bottom predictive accuracy are 1.0 and 0, respectively. The best models for evolutionary data [(B) to (D)] only use 3 predictors; we include a bar for affectiveness at 0 for illustration purposes.

creativity. The first includes 254 cases of overextension reported in English-speaking children (*15*), the most comprehensive collection available. We focused on English because overextension data from other languages is sparse and not suitable for a scalable analysis. The second data set draws on the Database of Cross-Linguistic Colexifications (CLICS) (*36*), the largest resource for colexification. We worked with 22,379 attested colexification cases from 1486 languages. Accompanying CLICS on the longer evolutionary timescale, we also analyzed a third dataset, DatSemShift (*37*). This is the largest resource of historical semantic change, covering 1792 attested cases of semantic change from 516 languages.

For modeling purposes, we balanced the data to include an equal number of positive and negative cases. Positive cases exhausted the attested pairs in each dataset after preprocessing. We randomly sampled negative cases from pairings of attested meanings that result in unattested combinations. Following this procedure, the task of the models is to use one or multiple semantic knowledge types (visual similarity, associativity, taxonomic similarity, or affectiveness) to characterize each phenomenon by contrast to a backdrop of negative cases (*10*, *11*, *15*).

To test our proposal, we first identified the model that best characterizes each of the three phenomena in isolation and then tested each of these models on data from the other phenomena. For model selection, we fit several logistic regression models predicting whether a pair of meanings colexifies in a language, participates in semantic change in a language, or appears in overextension in English. For each phenomenon, the only parameters that vary across their models are the knowledge types that they have as predictors. We considered all possible combinations: from four univariate models per phenomenon (one for each knowledge type) to models with two, three, or all four knowledge types. Colexification and semantic change models have language and geographical region as population-level effects. For model comparison and validation, we used approximate leave-one-out cross-validation. Our measure for model selection is expected log predictive density (*38*).

Results

Types of semantic knowledge

Figure 2, A to C, shows the standardized estimates from the best model for each phenomenon. Details and model rankings are provided in the supplementary materials. These results generalize previous findings that analyze each phenomenon separately (*3*, *10*, *11*, *15*) in several ways. First, the results show that across phenomena, a word is more likely to be creatively extended when its meaning shares properties with that of a new target referent. All coefficients are positive, meaning that the higher the semantic relatedness between two meanings, the higher the likelihood that they will be connected through lexical creativity. When meanings are similar along several dimensions, the likelihood that lexical creativity connects them grows. Second, the results suggest that the semantic properties that anchor lexical creativity are of diverse types. All of the best models draw on multiple knowledge types: for developmental data (Fig. 2A), all four of them, and for evolutionary data (Fig. 2, B and C), all modalities but affectiveness. A third finding is the similarity in the ranking of the coefficients, with associativity being the highest.

There are also differences between the models (Fig. 2, A to C). The most salient one is that associativity is more predictive of colexification and semantic change than of overextension. Taxonomy and visual similarity are more predictive of overextension than of the other two phenomena, and overextension factors in affectiveness, unlike the other two. These differences are partly reflected in the literature on child overextension, in which a prominently documented type of overextensional error is violation of taxonomic constraints—e.g., the use of words to describe referents from higher-order taxonomic categories (*6*). However, these differences might also be attributed to the resources we use as proxies for knowledge types. These resources are based on adult language use and may thus account less well for children's data [e.g., in word association, adult speakers are more likely to associate concepts based on situation as opposed to taxonomy (*39*)].

Cross-prediction

We next show that lexical creativity builds on a common foundation more directly, by performing a cross-predictive analysis. Specifically, we evaluated models for one phenomenon (e.g., overextension) on how well they account for data from a different phenomenon (e.g., colexification). We termed this "cross-prediction," and contrast it with "self-prediction," that is, prediction for unseen data from the same phenomenon (e.g., the overextension model being applied to unseen overextension data). To rule out any carryover owing to pairs that appear in multiple datasets [e.g., "moon" and "sun," a meaning pair that colexifies in some languages and children have linked through overextension (the intersection is illustrated in Fig. 1A)], we excluded all pairs that appear in more than one dataset. This makes the task harder but ensures that the models' performance reflects their capabilities to characterize truly out-of-sample data. Details can be found in the supplementary materials.

Figure 2D reports self- and cross-predictive accuracies of the models. Cross-prediction is very successful, even when compared with self-prediction. There is good carryover not only among the longer timescale phenomena but also between developmental and evolutionary phenomena. In all cases, the difference in accuracy between self- and cross-prediction is very small (between 0 and 0.03), and the difference to the baseline is large (0.22 to 0.31).

Because the models differ in their coefficients (Fig. 2, A to C), it could be that the self- and cross-predictions yield similar accuracies but make quite different predictions. This is not the case (Fig. 3): Self- and cross-predictions are well aligned throughout.

Robustness checks

Our study builds on English resources to derive proxies for semantic knowledge owing to a lack of comparable large-scale resources in other languages. This may not fully capture variation that is culture or language specific (*17*). Moreover, building on English resources could introduce English-specific biases; for example, a bias could be introduced because the overextension data are from English-speaking children. To ensure that such biases do not drive our findings, we performed a series of robustness checks (supplementary materials).

First, we reevaluated the models' cross-predictive abilities on data that exclude Indo-European languages, the language family to which English belongs. This exclusion only concerns colexification and semantic change data because the overextension data are based on English speakers. If cross-prediction results were driven by an English bias, we would expect the models' predictive capabilities to decrease when tested on non–Indo-European data only. In this regard, our results are robust (Fig. 4A).

Next, we performed checks by manipulating the data on which the models are fit. We refit the best colexification and semantic change models, leaving out one at a time each of the five major language families found within the data.

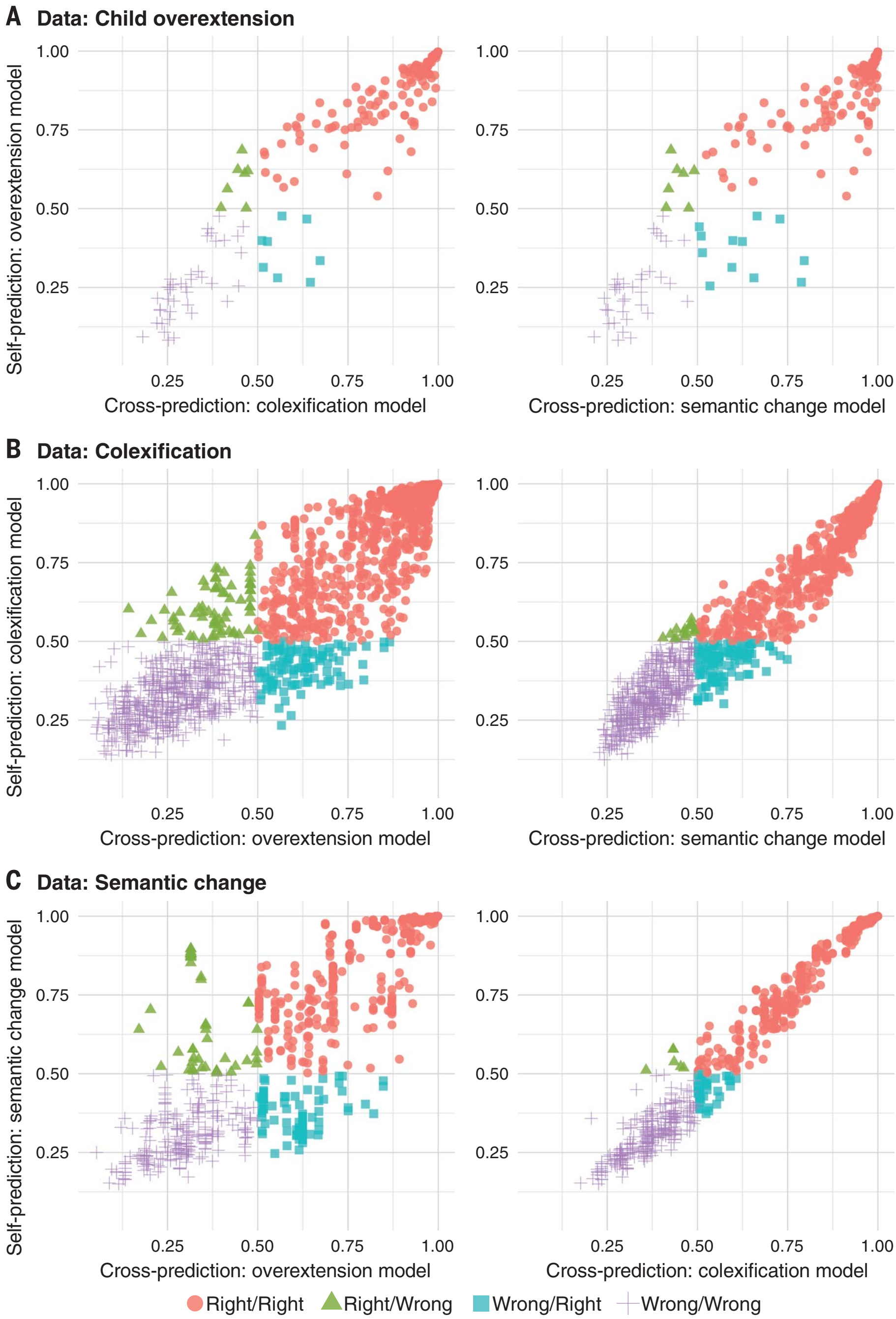

Fig. 3. Comparisons of model self- and cross-predictions based on data from child overextension, colexification, and semantic change. (**A** to **C**) Comparsions of self-predictions made by the best model of a phenomenon (*y* axis) against cross-predictions made by the best model for another phenomenon (*x* axis) for data from child overextension (A), colexification (B), and semantic change (C). Data points are attested cases of each phenomenon (a counterpart for unattested cases can be seen in fig. S3). Colors and shapes separate predictions into classes: "Right/Right" indicates correct predictions by both the self- and cross-predicting models, "Wrong/Wrong" indicates incorrect predictions by both, "Right/Wrong" indicates a correct prediction from the self-predicting model but an incorrect one from the cross-predicting one, and conversely for "Wrong/Right." To make plots legible, colexification data were randomly subsampled to 8%.

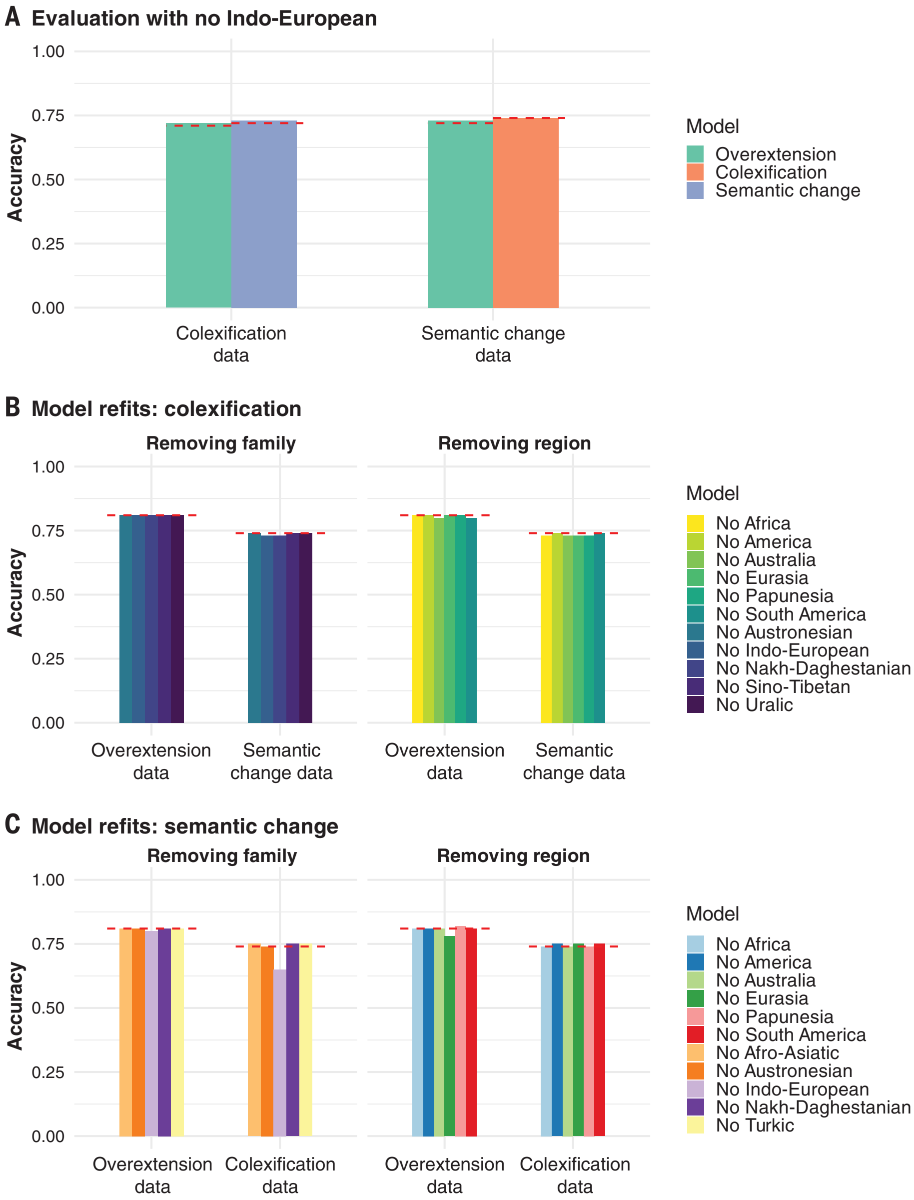

Fig. 4. Robustness checks on cross-prediction against geographic or phylogenetic biases. Red dashed lines indicate accuracy of best models on original data (Fig. 2D). (**A**) Accuracy of best models when evaluated on data that excludes Indo-European languages. (**B** and **C**) Accuracy of colexification (B) and semantic change (C) models when refit either excluding data from one of their five largest language families (left) or without data from one of their six macro regions (right).

The same leave-one-out refitting process was conducted for all large geographical regions. The results are stable (Fig. 4, B and C)

Last, we also redid our analyses using alternative visual representations obtained from a model trained on a nonlinguistic task. When we used these representations as well, our results were stable (tables S15 to S20 and figure S1).

Discussion

Our findings suggest a shared human capacity to creatively extend words to new meanings across timescales and the individual and population level. We argue that this capacity relies on a common foundation of knowledge, with different facets of semantic relatedness that enable new meaning extensions.

Although our results indicate that diverse manifestations of lexical creativity are related and share common ground, the current study cannot speak directly to the nature of this relationship. Our findings are compatible with at least two different explanations. The first explanation is a direct causal pathway, with child overextensions being adopted directly by linguistic communities, hence explaining their resemblance to products of language evolution. However, we believe that this account is implausible for several reasons. First, it is unlikely for children's spontaneous innovations to be regularly adopted by broader adult populations or in language change (*23*, *40*), and more so to a degree that leaves a cross-linguistic signature. Second, this account would leave the attested nonintersecting cases of colexification and semantic change in Fig. 1A unexplained. Some of these meanings are encountered relatively late in language acquisition, making them less likely to appear in overextension. Third, functional pressures toward efficient communication shape word meanings across languages (*11*, *41*–*44*). These pressures are independent of child overextension and suggest that phenomena such as colexification are partially shaped by a need to distinguish meanings that appear in similar contexts. This may explain why child overextensions such as "baby" for "adult" or "bus" for "train" are rarely expressed by a single word across languages: Doing so may cause ambiguity that is hard to resolve even in context (*11*, *44*, *45*). A second explanation, which we suggest to be more likely, is an indirect relationship, in which products of lexical creativity stem from a common latent source of multifaceted semantic knowledge (Fig. 1B). That is, children draw from this source for overextension, and adults do so as well when extending meaning in new ways. Some instances of creative lexical uses by adults (e.g., the metaphorical extension of "mouse" to the computer device) are then adopted by their linguistic communities over time, making their way into the lexicon.

We have shown that the products of lexical creativity of young learners and language users in general can both be explained by a single latent common ground. Our work identifies a foundation of shared knowledge for this common ground, extending prior research that suggests that words tend to express related meanings (*10*, *11*) owing to cognitive advantages for learning, retrieving, and interpreting words (*3*, *46*–*48*). Additionally, more generally, the use of the same word for multiple meanings allows for more compressible lexicons (*49*) and for the reuse of shorter words that are easier to produce (*45*, *50*). Future work should further specify the origins of this common foundation and the cognitive mechanisms of human lexical creativity.

REFERENCES AND NOTES

1. M. Bréal, *Essai de Sémantique: Science des Significations* (Hachette, 1897).
2. S. Ullmann, *Word* **9**, 225–240 (1953).
3. C. Ramiro, M. Srinivasan, B. C. Malt, Y. Xu, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 2323–2328 (2018).
4. L. S. Vygotsky, *Language and Thought* (MIT Press, 1962).
5. E. V. Clark, *Child Dev.* **49**, 953–959 (1978).
6. L. A. Rescorla, *J. Child Lang.* **7**, 321–335 (1980).

7. E. V. Clark, *Cognitive Development and Acquisition of Language*, T. E. Moore, ed. (Academic Press, 1973), pp. 65–110.
8. A. François, *Studies in Language Companion Series* (JB, 2008), pp. 163–215.
9. A. Majid, J. S. Boster, M. Bowerman, *Cognition* **109**, 235–250 (2008).
10. Y. Xu, K. Duong, B. C. Malt, S. Jiang, M. Srinivasan, *Cognition* **201**, 104280 (2020).
11. T. Brochhagen, G. Boleda, *Cognition* **226**, 105179 (2022).
12. E. C. Traugott, R. B. Dasher, *Regularity in Semantic Change* (Cambridge Univ. Press, 2001).
13. E. Sweetser, *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure* (Cambridge Univ. Press, 1990).
14. D. P. Wilkins, *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*, M. Durie, M. Ross, Eds. (Oxford Univ. Press, 1996), pp. 264–304.
15. R. Ferreira Pinto Jr., Y. Xu, *Cognition* **206**, 104472 (2021).
16. H. Youn *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **113**, 1766–1771 (2016).
17. J. C. Jackson *et al*., *Science* **366**, 1517–1522 (2019).
18. A. Di Natale, M. Pellert, D. Garcia, *Affect. Sci.* **2**, 99–111 (2021).
19. G. Stern, *Meaning and Change of Meaning with Special Reference to the English Language* (Indiana Univ. Press, 1931).
20. L. Bloomfield, *Language* (Allen & Unwin, 1933).
21. D. Bickerton, *Language and Species* (Univ. of Chicago Press, 1990).
22. T. Givón, *Evol. Commun.* **2**, 45–116 (1998).
23. D. Slobin, *Biology and Knowledge Revisited* (Routledge, 2004), pp. 273–304.
24. J. Culbertson, P. Smolensky, G. Legendre, *Cognition* **122**, 306–329 (2012).
25. B. Beekhuizen, A. Fazly, S. Stevenson, *Proc. Ann. Conf. Cogn. Sci. Soc.* 170–175 (2014).
26. B. Beekhuizen, S. Stevenson, *Cogn. Sci.* **42**, 2699–2734 (2018).
27. S. Kirby, H. Cornish, K. Smith, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10681–10686 (2008).
28. P. Monaghan, *Cognition* **133**, 530–534 (2014).
29. S. De Deyne, D. J. Navarro, A. Perfors, M. Brysbaert, G. Storms, *Behav. Res. Methods* **51**, 987–1006 (2019).
30. S. De Deyne, D. J. Navarro, G. Collell, A. Perfors, *Cogn. Sci.* **45**, e12922 (2021).
31. E. Gualdoni, T. Brochhagen, A. Mädebach, G. Boleda, Woman or tennis player? Visual typicality and lexical frequency affect variation in object naming, in *Proceedings of the 44th Annual Meeting of the Cognitive Science Society*, J. Culbertson, A. Perfors, H. Rabagliati, V. Ramenzoni, Eds., 2022.
32. P. Anderson *et al*., Bottom-up and top-down attention for image captioning and visual question answering, *IEEE/CVF Computer Vision and Pattern Recognition Conference*, 2018.
33. C. Fellbaum, *WordNet* (MIT Press, 2015).
34. A. B. Warriner, V. Kuperman, M. Brysbaert, *Behav. Res. Methods* **45**, 1191–1207 (2013).
35. S. Mohammad, Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words, *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics,* 2018.
36. C. Rzymski *et al*., *Sci. Data* **7**, 13 (2020).
37. A. Zalizniak, *et al.*, Database of semantic shifts. *Moscow: Institute of Linguistics, Russian Academy of Sciences* (2016–2020); https://datsemshift.ru/.
38. A. Vehtari, A. Gelman, J. Gabry, *Stat. Comput.* **27**, 1413–1432 (2017).
39. S. De Deyne, G. Storms, *Behav. Res. Methods* **40**, 213–231 (2008).
40. J. L. Bybee, D. I. Slobin, *Why Small Children Cannot Change Language on Their Own: Suggestions from the English Past Tense* (John Benjamins, 1982).
41. C. Kemp, T. Regier, *Science* **336**, 1049–1054 (2012).
42. C. Kemp, Y. Xu, T. Regier, *Annu. Rev. Linguist.* **4**, 109–128 (2018).
43. Y. Xu, E. Liu, T. Regier, *Open Mind (Camb.)* **4**, 57–70 (2020).
44. A. Karjus, R. A. Blythe, S. Kirby, T. Wang, K. Smith, *Cogn. Sci.* **45**, e13035 (2021).
45. S. T. Piantadosi, H. Tily, E. Gibson, *Cognition* **122**, 280–291 (2012).
46. D. Gentner, M. Bowerman, *Crosslinguistic Approaches to the Psychology of Language: Research in the Tradition of Dan Isaac Slobin*, J. Guo, *et al.*, Eds. (Psychology Press, 2009), pp. 465–480.
47. J. M. Rodd *et al*., *Mem. Cognit.* **40**, 1095–1108 (2012).
48. M. Srinivasan, C. Berner, H. Rabagliati, *J. Exp. Psychol. Gen.* **148**, 926–942 (2019).
49. Y. Xu, B. C. Malt, M. Srinivasan, *Cognit. Psychol.* **96**, 41–53 (2017).
50. G. Zipf, *The Psycho-Biology of Language* (Houghton Mifflin, 1935).
51. T. Brochhagen, G. Boleda, E. Gualdoni, Y. Xu, From language development to language evolution: A unified view of human lexical creativity, OSF (2023); https://doi.org/10.17605/OSF.IO/ZKGU3.

ACKNOWLEDGMENTS

We are grateful for the feedback we received from M. Baroni, D. Dediu, C. Kemp, B. C. Malt, L. McNally, S. Padó, T. Regier, M. Srinivasan, S. Stevenson, and L. Yu. We also thank the reviewers for their helpful comments on our manuscript. **Funding:** G.B. and E.G. received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no 715154; AMORE). G.B., E.G., and T.B. are also funded by Ministerio de Ciencia e Innovación and the Agencia Estatal de Investigación (Spain; ref. PID2020-112602GB-I00/MICIN/AEI/10.13039/501100011033). Y.X. is funded by NSERC Alliance International Catalyst grant ALLRP 576149-22, NSERC Discovery grant RGPIN-2018-05872, SSHRC Insight grant 435190272, Ontario Early Researcher award ER19-15-050, and an Amazon Research award. No funding agency was involved in the conceptualization, design, data collection, analysis, decision to publish, or preparation of this manuscript, and the views expressed in this manuscript do not necessarily reflect the views of our funding agencies. **Author contributions:** T.B., G.B., and Y.X. conceptualized and designed the study. T.B. and E.G. acquired and analyzed the data. T.B., G.B., and Y.X. interpreted the analysis. G.B. and Y.X. acquired funding. T.B., G.B., and Y.X. wrote the manuscript. All authors approved the submitted manuscript. **Competing interests:** The authors have no competing interests to declare. **Data and material availability:** All code and material for replicating our analyses can be found at (*51*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade7981
Supplementary Text
Materials and Methods
Figs. S1 to S3
Tables S1 to S22
References (*52*–*66*)
MDAR Reproducibility Checklist

Submitted 27 October 2022; accepted 1 June 2023
10.1126/science.ade7981