Making room for new words

Claire Augusta Bergey*
cbergey@stanford.edu
Linguistics, Stanford University

Aja Altenhof*
altenhof@wisc.edu
Linguistics, University of Wisconsin-Madison

Charlotte Burnham
cvb@stanford.edu
Linguistics, Stanford University

Oliver Lee
oliver.lee@stanford.edu
Symbolic Systems, Stanford University

Robert D. Hawkins
rdhawkins@stanford.edu
Linguistics, Stanford University

Abstract

When words are borrowed into a language, their meanings often narrow. For example, *salsa* can mean any sauce in Spanish but refers only to tomato-pepper sauces in English. What principles of individuals' learning shape this collective semantic narrowing? We propose that semantic narrowing reflects an accumulation of pragmatic inferences about word meaning. When listeners encounter a novel word for an already-nameable concept, they infer that it has a more specific and atypical meaning than existing terms. Across four experiments, we tested these inferences among adults and preschoolers. Participants saw arrays of familiar category exemplars (e.g., leaves of varying subtypes) and were prompted with familiar words ("leaves") or novel words ("nefts"). Both adults and children selected more specific subtypes and more atypical exemplars when prompted with novel words. Learners carve out distinctive meanings for new words, a pragmatic inference that may shape meaning from early childhood to long-range language change.

Keywords: linguistics, language acquisition, pragmatics, concepts and categories, cognitive development

Introduction

What happens when a word is borrowed from one language into another? When the word's meaning is similar to word meanings in the recipient language, borrowers seem to avoid taking on synonymous meanings—and in systematic ways. In some cases, the meaning of the borrowed word narrows. In Polish, the borrowed English word *dog* refers not to dogs in general, but specifically to Great Danes (Fisiak, 1970). In English, the borrowed Hindi word *chai* (lit. "tea") refers specifically to spiced, milky tea. These cases of semantic narrowing are not isolated. A distributional analysis of Anglicisms in a large corpus of Argentinian newspapers revealed that loanwords display less contextual variability than their native competitors (Serigos, 2017). 70% of English loanwords on Serbian social media undergo some kind of semantic narrowing, often only retaining a single sense from polysemous source words (Vuksanović, 2019). That is, loanwords with an existing semantic counterpart seem to require differentiation to justify their presence in the lexicon (Onysko & Winter-Froemel, 2011).

What cognitive mechanisms underpin semantic narrowing in language contact? Long-range language change must be instantiated through individual instances of language use. Lab experiments can help establish plausible individual-level cognitive mechanisms underlying language change (Kirby, Cornish, & Smith, 2008; Smith, Kirby, & Brighton, 2003). We propose that semantic narrowing may be underpinned by the accumulation of individual-level contrastive inferences: when speakers encounter a new word for an already nameable concept, they assume it must refer to something more specific. Language documentation and experimental tasks show that people avoid absolute synonymy (Altenhof & Roberts, 2023; Andersen, 2014; Cruse, 1986; Geeraerts, 2009). These phenomena undergird the Principle of Contrast: people avoid having multiple language forms that correspond to the exact same meanings (Clark, 1990, 2014).

The Principle of Contrast states that speakers avoid exact synonymy, but its implications for how learners accommodate new words into an already-populated semantic space have not been fully specified. Here, we extend the Principle of Contrast to propose a pragmatic constraint on the lexicon: language users aim to both maximize coverage of their conceptual space and to prevent multiple concepts from exactly overlapping. That is, they aim for a neatly tiled conceptual space in which every distinction between words makes a conceptual difference, preferring to use new words to cover gaps, subordinate regions, and atypical areas of conceptual space. We call this the *conceptual tiling* account of meaning accommodation.

The process of accommodating word meanings into an already populated conceptual space is also the process of word learning generally. To the extent that tiling constraints are present in early childhood, they may shape how children infer word meanings as they acquire language. Word learning research has established several principles by which children accommodate new meanings. When presented with a novel object and a familiar one, children assume that an unfamiliar label applies to the novel object (*mutual exclusivity*) (Markman & Wachtel, 1988; Waxman, Philippe, & Branning, 1999). This tendency has been explained by some researchers as a simple bias—such as an assumption that any given object has only one category label (Markman & Wachtel, 1988) or a tendency to apply novel words to lexical gaps (Markman, Wasow, & Hansen, 2003)—and by others as deriving from a general pragmatic constraint, the Principle of Contrast (Clark, 1990, 2014). Importantly, pragmatic accounts specify *not* that a given referent has only one label, but that each concept must be distinct from other concepts in the extent of meanings it encompasses. While pragmatic and bias-based theories make the same predictions about tasks with novel objects (that peo-

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

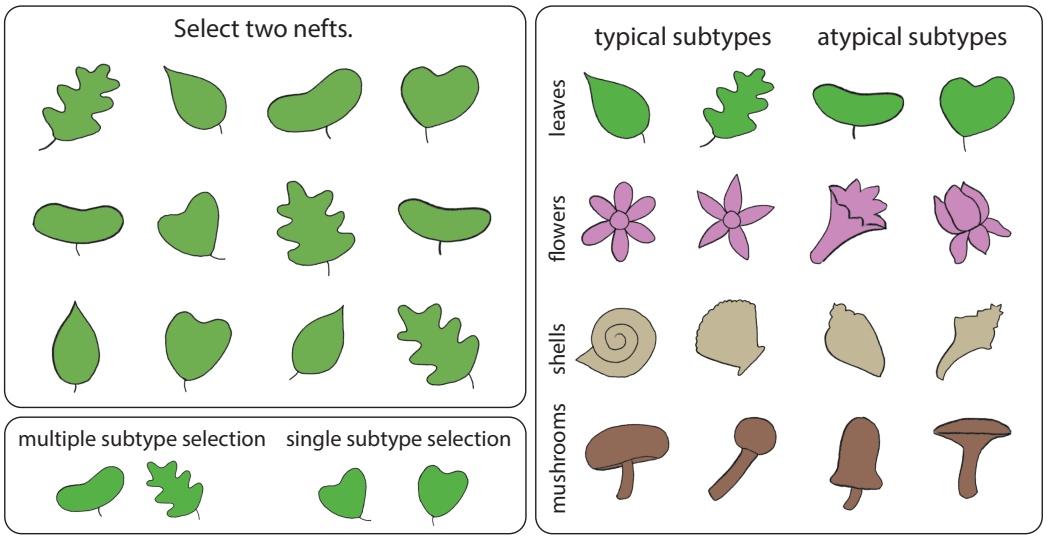

Figure 1: Stimuli in Experiments 1 and 2. Left: An example stimulus display and prompt; children would be prompted via audio with "Tap on two nefts." Below are examples of a multiple subtype selection and a single subtype selection. Right: Examples of all subtypes in the experiments, divided into typical and atypical subtypes.

ple tend to fill gaps), only pragmatic theories make systematic predictions about how people accommodate their conceptual spaces in the face of novel labels for familiar objects. We therefore introduce a task that involves only familiar objects to test the first distinctive predictions of the conceptual tiling account.

Our task design takes advantage of several other established principles of word learning. All else equal, adults and children tend to assume that words refer to basic-level categories—categories that tend to share a general shape and function and have a simple single-word name (e.g., *leaf, dog, apple, shoe*) (Golinkoff, Mervis, & Hirsh-Pasek, 1994; Markman, 1989; Rosch, 1975). How, then, do they learn words at the subordinate—more specific than basic—level? Children and adults assume that speakers sample across the breadth of a category (*strong sampling*), such that when shown three dalmatians labeled with a novel word, they infer that the word applies to a subordinate category (dalmatians), not the basic-level one (dogs) (Xu & Tenenbaum, 2007). Adults and children also assume that words applied to atypical exemplars (e.g., a blowfish compared to a typical goldfish) refer to subordinate categories (Emberson, Loncar, Mazzei, Treves, & Goldberg, 2019). Here, we take advantage of people's tendencies to sample across the breadth of a category and associate atypical exemplars with subordinate categories in our task.

We test the conceptual tiling account across four controlled concept learning experiments with adults and 3- to 5-year-old children. Preschool-aged children were selected because the 3- to 5-year-old age range is a time of rapid word learning, a time when other word learning biases discussed above are in place, and a time when children have a grasp on many basic-level terms to accommodate new meanings around. Participants were shown arrays of familiar objects (e.g., leaves) and asked to select two of them (e.g., "Tap on two [leaves/nefts]"). By asking for two selections, we can examine the breadth of the categories they choose, exploiting the fact that people should tend to sample across the extent of their inferred category. If people apply a conceptual tiling constraint, inferring that novel words applied to familiar referents must have distinct and subordinate meanings, they will choose more specific and atypical sets of exemplars when prompted with novel words compared to familiar words. This tendency, if present, would shape children's accommodation of word meaning across development, adults' accommodation of new vocabulary, and, over the course of many individual-level inferences, the semantic narrowing of borrowed words.

Experiments 1 & 2

Method

Participants. In Experiment 1[1], adult participants ($N = 61$) were recruited through Prolific for a custom jsPsych web experiment (De Leeuw, Gilbert, & Luchterhandt, 2023) and paid $1.00 for participation. Adult participants were screened for being located in the US, UK, or Canada and speaking English as a first language.

In Experiment 2, child participants ($N = 48$; 18 3-year-olds, 18 4-year-olds, and 12 5-year-olds) were recruited from English-speaking classrooms at a US preschool to do a custom jsPsych tablet-based task.

[1]Experiment code, analysis code, and data for all experiments are available at https://github.com/aaltenhof/synonymous_borrowing.

Procedure. Participants were assigned to either the familiar word condition or the novel word condition. On each trial, participants saw a grid of 12 items of the same basic-level kind (e.g., leaves). These items were divided equally among four subtypes (e.g., oak-shaped leaves, teardrop-shaped leaves, heart-shaped leaves, and bean-shaped leaves). Two subtypes were designated as typical by experimenters (e.g., oak-shaped and teardrop-shaped leaves), and two subtypes were designated as atypical (e.g., heart-shaped and bean-shaped leaves). Individual exemplars of each subtype varied slightly in shape and orientation. We designed all stimuli specifically for this task (Figure 1).

Participants were prompted to select two items using either a familiar or novel word, adults via text (e.g., "Select two [leaves/nefts]") and children via audio prompt (e.g., "Tap on two [leaves/nefts]"). By asking participants to select two items, we are able to evaluate how specific they judged the word meanings to be (i.e., two items of one subtype indicates a more specific category; two items of different subtypes indicates a broader category). Participants tapped or clicked on the images to respond.

Familiar words were basic-level terms (leaves, flowers, mushrooms, and shells) determined to be known to most 4-year-olds (ages of acquisition: 21.6 months, 22 months, 42.6 months, and 49.5 months respectively; norms derived from Frank, Braginsky, Yurovsky, and Marchman (2017) or where not available from Morrison, Chappell, and Ellis (1997)). Novel words were designed to have English phonotactics (nefts, palts, tinches, and bines). Each participant carried out four trials, one of each category (leaves, flowers, mushrooms, and shells) in random order. Assignment of participants to novel or familiar word condition, novel words to categories, and screen position of items were all randomized.

Children completed two initial practice trials with familiar categories and words: one display of apples and strawberries in which they were asked to tap on two apples and one array of carrots and broccoli in which they were asked to tap on two carrots. Practice trial displays were evenly split between two basic-level categories so that children would not be clued into the structure of experiment trials, and were intended to ensure children could reliably select two items of a requested kind. A child's data were excluded from analysis if the child saw the stimuli multiple times due to experimenter or technical errors (excluding n = 1 mistaken re-recruitment of the same child), or the child did not make four correct selections (100% accuracy) on practice trials (n = 5), leaving a final N = 43 (15 3-year-olds, 16 4-year-olds, and 12 5-year-olds). Adults were not included or counted toward total recruitment if they did not complete the task (n = 2), leaving a final $N = 61$.

Stimulus norming. To verify experimenter-designated typicality values, image stimuli were normed for typicality post-hoc in a separate judgment task among adults. We recruited 31 participants on Prolific using the same inclusion criteria as above and excluding anyone who participated in Experiment 1 or Experiment 3; participants were paid $1.50 for their time.

On each trial, participants judged one image stimulus on the prompt "How typical is this object for a [mushroom]?" on a slider scale from "Very atypical (unusual, rare, abnormal)" to "Very typical (usual, common, normal)". Each participant judged one randomly selected item from each subtype in each category (e.g., one teardrop-shaped leaf, one bean-shaped leaf, and so on) and two items from each child practice trial category (e.g., two strawberries) for a total of 40 trials in random order. Note that stimuli for Experiments 1, 2, 3, and 4 were all rated in this task; we report norms for stimuli in Experiments 1 and 2 here.

Slider selection values (0-100) were analyzed to test whether experimenters' typicality designations were reflected among these independent raters. As intended, high-typicality items were rated significantly higher than low-typicality items ($M = 72.56$ compared to $M = 41.22$, $p < .001$ by t-test). Tests comparing typicality disparities within each category (leaves, flowers, shells, and mushrooms) were all significant (all ps $< .001$). Note that low-typicality items did not tend to fall extremely on the "Very atypical" side of the scale (range of low typicality item means: 10.40 to 58.10, $M = 41.22$); that is, we are contrasting high-typicality with moderately atypical items, and any findings do not rely on extreme atypicality. As intended, all child practice trial items were highly typical (range of item means: 71.55 to 94.40). These results verify that the items effectively manipulated typicality.

Results

We predicted that participants would choose (1) more specific selections (i.e., two items of the same subtype) and (2) more atypical items when prompted with novel words compared to familiar words. To evaluate these hypotheses, we model the number of distinct subtypes chosen on each trial (coded as 0 for two of the same subtype, 1 for two distinct subtypes) and the number of atypical items (0, 1, or 2) chosen on each trial as a function of word condition (familiar vs. novel, effect coded as 1, -1). We use generalized linear mixed-effects models with a binomial link; all models include by-subject and by-category random effects, and the child models include a mean-centered age predictor and its interaction with word condition.

Adults choose more atypical items when prompted with a novel word compared to a familiar word (26% for familiar vs. 60% for novel, $\beta = -0.81, p < .001$). Adults also choose more specific selections (i.e., are less likely to choose items of different subtypes) when prompted with a novel word compared to a familiar word (57% choosing two subtypes for familiar vs. 24% for novel, $\beta = 1.15, p < .001$). That is, taking the two measures together, people tend to choose distinct, typical subtypes of a category when prompted with a familiar word. By contrast, when prompted with a novel word, they tend to choose multiple exemplars of the same atypical subtype (Figure 2).

Do children also show evidence of semantic narrowing? We find that children choose more atypical items when prompted with a novel compared to a familiar word (36%

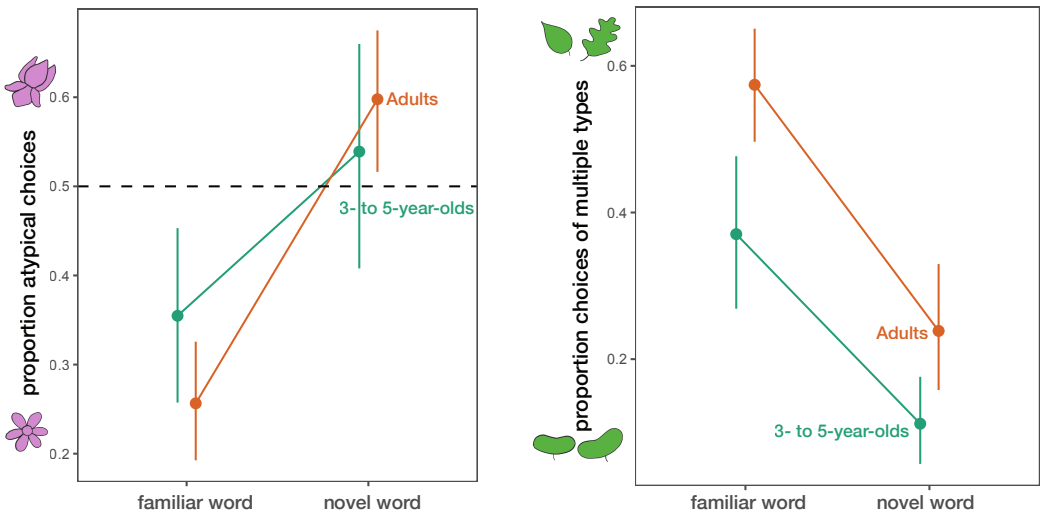

Figure 2: Results from Experiments 1 and 2. Adults and 3- to 5-year-old children choose more atypical and more specific selections of exemplars when prompted with novel words. Dashed line represents chance selection of atypical items; chance selection of multiple types would be 82% (after choosing any random first item, there is a 2/11 or 18.18% chance of choosing another of the same kind), off the axis range.

atypical for familiar vs. 54% for novel, β = -0.48, $p = .025$). Further, children choose more specific sets of referents when prompted with a novel compared to a familiar word (37% choosing multiple subtypes for familiar vs. 11% for novel, β = 1.04, $p = .001$) (Figure 2). We do not find effects of age or interactions with age in our models of typicality (β_{age} = -0.00, $p = .944$; $\beta_{word*age}$ = -0.03, $p = .161$) or number of subtypes (β_{age} = -0.04, $p = .216$; $\beta_{word*age}$ = -0.00, $p = .992$). In sum, much like adult participants, children tend to select across multiple typical subtypes when prompted with familiar words, but when prompted with novel words they tend to select multiple exemplars of the same atypical subtype (Figure 2).

We find that both adults and 3- to 5-year-old children show robust evidence of semantic narrowing: they assume that novel words have narrower and more atypical meanings than the words already in their lexicon.

Experiments 3 & 4

To strengthen the findings of Experiments 1 and 2, we pre-registered the method, analyses, and hypotheses of two replication experiments with larger sample sizes, more trials per participant, and an expanded stimulus set. Links to pre-registrations are included in the GitHub repository designated above. We broaden the stimulus set to include both natural categories and artifact categories. Natural categories are often treated as having underlying essences (Gelman & Markman, 1987; Kalish, 1995) and can display more graded category membership (Diesendruck & Gelman, 1999), while artifact categories are often treated as having necessary shape and function features (Jones & Smith, 2002). We do not have particular predictions about differences (or lack thereof) in meaning accommodation across natural and artifact categories, but rather aim to determine whether this inference holds across different classes of stimuli. In sum, these experiments give us better statistical power by which to establish these findings and the ability to evaluate stimulus generalization across domains.

Method

Participants. In Experiment 3, a pre-registered sample of $N = 120$ adult participants were recruited through Prolific for a web experiment with the same recruitment parameters as Experiment 1 and without re-recruiting individuals who participated in Experiment 1.

In Experiment 4, we pre-registered a sample of 80 child participants and here analyze the data of $N = 37$ child participants collected thus far (15 3-year-olds, 18 4-year-olds, and 3 5-year-olds) who were recruited from the same preschool setting as Experiment 2 without re-recruiting children who participated in Experiment 2.

As pre-registered, a child's data were excluded from analysis if the child saw the stimuli multiple times due to experimenter or technical errors (n = 5) or did not make four correct selections (100% accuracy) on practice trials (n = 7). One child's data was lost due to a technical error. This results in a sample of 26 children (9 3-year-olds, 15 4-year-olds, and 2 5-year-olds). Adult participants who did not complete the task or pass an included CAPTCHA were not included or counted toward recruitment (n = 10), and a few participants' data were not recorded due to a technical error (n = 3), for a final sample of N = 117.

Procedure. Tasks were nearly identical to Experiments 1 and 2, with an expanded stimulus set. Each participant com-

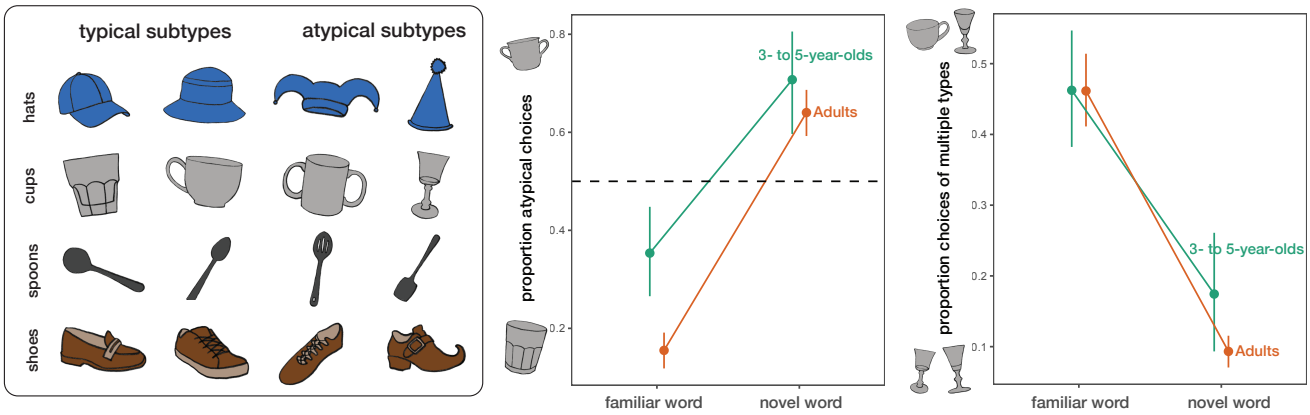

Figure 3: Left: Example artifact stimuli from Experiments 3 and 4. Right: Results from Experiments 3 and 4. Adults and 3- to 5-year-old children choose more atypical and more specific selections of exemplars when prompted with novel words, across natural and artificial categories. Dashed line represents chance selection of atypical items; chance selection of multiple types would be 82% (after choosing any random first item, there is a 2/11 or 18.18% chance of choosing another of the same kind), off the axis range.

pleted eight trials, equally divided between natural categories (leaves, flowers, mushrooms, and shells) and artifact categories (shoes, hats, spoons, and cups) and in random order (Figure 3, left panel). The artifact categories were chosen to be familiar to preschoolers (ages of acquisition: 17 months, 19 months, 21 months, and 21 months, respectively; norms derived from Frank et al. (2017)). Additional novel words (datch, sern, gade, and leam) were chosen to have English phonotactics.

Stimulus norming. Stimuli were normed in the typicality judgment experiment reported above; results for the artifact stimuli unique to Experiments 3 and 4 are reported here. As intended, high-typicality items were rated significantly higher than low-typicality items (M = 83.71 compared to M = 42.40, $p < .001$ by t-test). Tests comparing typicality disparities within each category (cups, hats, spoons, and shoes) were all significant (all ps < .001). Note again that low-typicality items did not tend to fall extremely on the "Very atypical" side of the scale (range of low typicality item means: 15.30 to 68.70, M = 42.40), such that our findings do not rely on extreme atypicality.

Results

As in Experiments 1 and 2, we model specificity of choices and typicality of choices as a function of word condition, adding an additional predictor of category type (natural vs. artifact, effect coded as (-1, 1)). Pre-registered model structures are otherwise identical.

Adults robustly choose more atypical items when prompted with a novel word compared to a familiar one (16% for familiar vs. 64% for novel, $\beta = -1.31, p < .001$). There is a marginal effect of stimulus domain such that people make slightly more atypical choices among artifacts ($\beta = 0.20, p = .078$). Adults also robustly choose more specific referents when prompted with a novel word compared to a familiar word (46% choosing two subtypes for familiar vs. 9% for novel, $\beta = 2.40$, $p < .001$). There is a small but significant effect of stimulus domain such that people more often choose multiple subtypes among natural categories ($\beta = -0.39, p = .029$). That is, we replicate that adults infer that novel words refer to more specific and more atypical categories than familiar words, and find that this effect generalizes across natural and artifact categories with only small differences between those domains (Figure 3).

Do children also display evidence of semantic narrowing? Even among this small initial subset of our pre-registered sample, we find that they do. Children choose more atypical items when prompted with a novel word compared to a familiar one (35% for familiar vs. 71% for novel, $\beta = -0.79, p < .001$). We find a small effect of age such that children choose more atypical items over development ($\beta_{age} = 0.04$, $p = .024$) and no significant interaction between age and word condition ($\beta_{word*age} = -0.02$, $p = .333$). We also find an effect of stimulus domain such that children choose more atypical items among artifact stimuli ($\beta = 0.33, p = .033$). Examining the number of subtypes chosen, children choose more specific sets of referents when prompted with a novel word compared to a familiar word (46% choosing multiple subtypes for familiar vs. 18% for novel, $\beta = 1.28, p < .001$). We do not find a significant effect of age, interaction with age, or effect of stimulus domain ($\beta_{age} = -0.04$, $p = .275$; $\beta_{word*age} = 0.06$, $p = .120$; $\beta_{stimulus} = -0.13$, $p = .676$) (Figure 3).

Do any small differences between natural and artifact categories reflect genuine inferential differences between those domains, or just that our stimuli show some incidental variation? We tested this question in an exploratory analysis using our elicited typicality judgments: if differences disappear

when our continuous typicality judgments are included as a predictor, this suggests that any differences are likely due to item-specific variation in typicality. We calculate the typicality differential for each stimulus category by taking the difference between the mean typicality judgments of its typical and atypical members. We use a generalized linear mixed-effects model with the same structure as our previous models and adding a predictor for typicality differential and its interaction with word condition. We find that among adults, typicality differential and its interaction with word condition significantly predict atypical choices ($\beta_{typicality} = -0.03, p = .019$; $\beta_{typicality*word} = -0.05, p < .001$) and the effect of stimulus domain remains marginally significant ($\beta = 0.28, p = .055$). Typicality differential does not significantly predict number of distinct subtypes chosen, nor does its interaction with word condition $\beta_{typicality} = -0.01, p = .691$; $\beta_{typicality*word} = -0.01, p = .535$) and the effect of stimulus domain remains marginally significant ($\beta = -0.34, p = .084$). No effects of typicality differentials or stimulus domain remain significant among children when adding this predictor into the model. In sum, item-wise variation in typicality likely account for effects of stimulus domain in part, though further research could more precisely isolate these factors.

Discussion

A new word enters an already crowded conceptual space. Where will it fit in? Our experiments demonstrate that adults and preschool-aged children assume that new words have subordinate meanings to familiar words. Across four experiments, adults and children showed consistent inferences that new categories did not have exact overlap with existing categories, instead preferring to assign them more specific and more atypical meanings. This inference is consistent with an account positing that learners aim for neatly tiled semantic spaces: no two terms should be exactly coextensive in meaning.

In this paper, we focus on the case of semantic narrowing, a commonly attested phenomenon in the literature on borrowing. However, conceptual spaces may shift in several ways in order to accommodate new meanings. In the simplest case, a loanword might fill a lexical gap in the recipient language (e.g., English *sushi* from Japanese), no accommodation necessary. When the borrowed word's meaning is coextensive with an existing word's meaning, speakers have several options. First, they may go subordinate, the narrowing we explore here. Second, they may go superordinate: for example, *armada*, from Spanish meaning a 'fleet of warships,' was borrowed into Russian as армáда denoting any large set of powerful units—not only ships, but also airplanes, tanks, or crowds of people (Bakurova, 2025). Third, a loanword may unseat an existing word, either doing away with it or forcing it into a subordinate position. For example, 桃色 (*momoiro*) was the default basic color term for 'pink' in Japanese from at least 750 CE to the late 1900s, when it was displaced by the English loan ピンク (*pinku*); *momoiro* has since slightly narrowed in meaning to encompass only lighter pinks (Stanlaw, 2010).

Why might narrowing be a particularly common strategy? People may hold a general assumption that they have learned most concepts at the superordinate and basic levels and will tend to learn more and more specific concepts over time. This aligns with research on language acquisition suggesting that basic-level terms are learned earliest, followed by superordinate terms, and finally by subordinate ones (Mervis & Crisafi, 1982). The conceptual tiling account, as a constraint on the entire lexicon, is consistent with all of the meaning accommodation strategies above but does not yet fully explain which factors push people to move a new word to a subordinate or superordinate level.

The behavior of individual loanwords is undoubtedly shaped by many factors in language contact, including relevance to the source culture (Winter-Froemel & Onysko, 2012; Zenner, Speelman, & Geeraerts, 2012), the source language's prestige (Bullock, Serigos, & Toribio, 2016), and the degree of contact between languages (Zenner, Speelman, & Geeraerts, 2013). Words may also differ in register but not conceptual coverage (e.g., English *urine, pee*, and *piss*), making the accommodation process particularly difficult to map when it spans high- and low-prestige registers. More abstract concepts are also more difficult to map in conceptual space than leaves and flowers. For example, the Finnish native politeness marker *kiitos* and its borrowed English analogue *pliis* have consistent differences in grammatical, pragmatic, and social usage (Peterson & Vaattovaara, 2014), but these usages are less easily mapped in a feature space. Though the concrete, imageable nouns we use here may present the simplest case to test, we expect the phenomenon of conceptual tiling to hold—and to require different measurement—across domains.

While we present an inference mechanism that plausibly underpins semantic narrowing, several steps remain in scaling up individuals' inferences to population-level usage. As each person's pragmatic inferences shape their understanding of a concept, their concept in turn shapes their downstream word use and provides evidence to subsequent learners (Bybee, 2010). Over many interactions, one-off pragmatic inferences can conventionalize into general-use semantics (Traugott & Dasher, 2001). Chains of transmission may even exaggerate initial narrowing inferences, a phenomenon that might be fruitfully investigated using iterated learning experiments (Tamariz & Kirby, 2016). On the level of the lexicon, these inferences give rise to informative conceptual distinctions, bolstering their stability (Clark, 1990; Traugott, 2012). The same conceptual tiling that shapes a child's word learning may thus, compounded across speakers and generations, shape the lexicon itself.

References

Altenhof, A., & Roberts, G. (2023). Quality, not quantity, impacts the differentiation of near-synonyms. *Language*

and Cognition, *15*(4), 854–883.

Andersen, G. (2014). Pragmatic borrowing. *Journal of Pragmatics*, *67*, 17–33.

Bakurova, M. V. (2025). Semantic adaptation of borrowings: A case study of Spanish loanwords in the Russian language. *Sibirskii Filologicheskii Zhurnal [Siberian Journal of Philology]*, *2*, 145–157.

Bullock, B. E., Serigos, J., & Toribio, A. J. (2016). The stratification of English-language lone-word and multi-word material in Puerto Rican Spanish-language press outlets. *Spanish-English Codeswitching in the Caribbean and the US*, *11*, 171.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press.

Clark, E. V. (1990). On the pragmatics of contrast. *Journal of child language*, *17*(2), 417–431.

Clark, E. V. (2014). The principle of contrast: A constraint on language acquisition. In *Mechanisms of language acquisition* (pp. 1–33). Psychology Press.

Cruse, D. A. (1986). *Lexical semantics*. Cambridge University Press.

De Leeuw, J. R., Gilbert, R. A., & Luchterhandt, B. (2023). jspsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *Journal of Open Source Software*, *8*(85), 5351.

Diesendruck, G., & Gelman, S. A. (1999). Domain differences in absolute judgments of category membership: Evidence for an essentialist account of categorization. *Psychonomic Bulletin & Review*, *6*(2), 338–346.

Emberson, L. L., Loncar, N., Mazzei, C., Treves, I., & Goldberg, A. E. (2019). The blowfish effect: children and adults use atypical exemplars to infer more narrow categories during word learning. *Journal of Child Language*, *46*(5), 938–954.

Fisiak, J. (1970). The semantics of English loanwords in Polish. *Studia Anglica Posnaniensia*, *2*, 41–49.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, *44*(3), 677–694.

Geeraerts, D. (2009). *Theories of lexical semantics*. Oxford University Press.

Gelman, S. A., & Markman, E. M. (1987). Young children's inductions from natural kinds: The role of categories and appearances. *Child Development*, 1532–1541.

Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: the case for a developmental lexical principles framework. *Journal of Child Language*, *21*(1), 125–155.

Jones, S. S., & Smith, L. B. (2002). How children know the relevant properties for generalizing object names. *Developmental Science*, *5*(2), 219–232.

Kalish, C. W. (1995). Essentialism and graded membership in animal and artifact categories. *Memory & Cognition*, *23*(3), 335–353.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681–10686.

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. mit Press.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121–157.

Markman, E. M., Wasow, J. L., & Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology*, *47*(3), 241–275.

Mervis, C. B., & Crisafi, M. A. (1982). Order of acquisition of subordinate-, basic-, and superordinate-level categories. *Child Development*, 258–266.

Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *The Quarterly Journal of Experimental Psychology Section A*, *50*(3), 528–559.

Onysko, A., & Winter-Froemel, E. (2011). Necessary loans–luxury loans? Exploring the pragmatic dimension of borrowing. *Journal of Pragmatics*, *43*(6), 1550–1567.

Peterson, E., & Vaattovaara, J. (2014). Kiitos and pliis: The relationship of native and borrowed politeness markers in Finnish. *Journal of Politeness Research*, *10*(2), 247–269.

Rosch, E. (1975). Universals and cultural specifics in human categorization. *Cross-cultural perspectives on learning*, *177206*.

Serigos, J. (2017). Using distributional semantics in loanword research: A concept-based approach to quantifying semantic specificity of Anglicisms in Spanish. *International Journal of Bilingualism*, *21*(5), 521–540.

Smith, K., Kirby, S., & Brighton, H. (2003). Iterated learning: A framework for the emergence of language. *Artificial life*, *9*(4), 371–386.

Stanlaw, J. (2010). Language, contact, and vantages: Fifteen hundred years of Japanese color terms. *Language sciences*, *32*(2), 196–224.

Tamariz, M., & Kirby, S. (2016). The cultural evolution of language. *Current Opinion in Psychology*, *8*, 37–43.

Traugott, E. C. (2012). Pragmatics and language change. *The Cambridge Handbook of Pragmatics*, 549–566.

Traugott, E. C., & Dasher, R. B. (2001). *Regularity in semantic change* (Vol. 97). Cambridge University Press.

Vuksanović, S. (2019). Semantic analysis of English loanwords in communication on the social networking website Facebook. Филолог – часопис за језик књижевност и културу. doi: 10.21618/fil1920198v

Waxman, S. R., Philippe, M., & Branning, A. (1999). A matter of time: Novel nouns mark object categories when delays are imposed. *Developmental Science*, *2*(1), 59–66.

Winter-Froemel, E., & Onysko, A. (2012). Proposing a pragmatic distinction for lexical Anglicisms. In *The angliciza-*

tion of european lexis (pp. 43–64). John Benjamins Publishing Company.

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*(2), 245.

Zenner, E., Speelman, D., & Geeraerts, D. (2012). Cognitive sociolinguistics meets loanword research: Measuring variation in the success of Anglicisms in Dutch. *Cognitive Linguistics*, *23*(4).

Zenner, E., Speelman, D., & Geeraerts, D. (2013). What makes a catchphrase catchy? Possible determinants in the borrowability of English catchphrases in Dutch. *New perspectives on lexical borrowing*, 41–64.