



**HAL**  
open science

# Word Learning: Homophony and the Distribution of Learning Exemplars

Isabelle Dautriche, Emmanuel Chemla, Anne Christophe

► **To cite this version:**

Isabelle Dautriche, Emmanuel Chemla, Anne Christophe. Word Learning: Homophony and the Distribution of Learning Exemplars. *Language Learning and Development*, 2016, 12 (3), pp.231-251. 10.1080/15475441.2015.1127163 . hal-02474121

**HAL Id: hal-02474121**

**<https://hal.science/hal-02474121>**

Submitted on 2 Feb 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Word learning: Homophony and the distribution of learning exemplars

Isabelle Dautriche\*, Emmanuel Chemla, and Anne Christophe

Laboratoire de Sciences Cognitives et Psycholinguistique (ENS, CNRS, EHESS), Ecole Normale Supérieure, PSL Research University, Paris, France

## Acknowledgements

This work was funded by grants from the European Research Council (FP/2007-2013-ERC n°313610), the Région Ile-de-France, the Fondation de France, LabEx IEC (ANR-10-LABX-0087), IdEx PSL (ANR-10-IDEX-0001-02), the ANR ‘Apprentissages’ (ANR-13- APPR-0012), and a PhD fellowship from Direction Générale de l’Armement (DGA). We thank Anne-Caroline Fievet and the language group of the LSCP.

## Abstract

How do children infer the meaning of a word? Current accounts of word learning assume that children expect a word to map onto exactly one concept whose members form a coherent category. If this assumption was strictly true, children should infer that a homophone, such as “bat”, refers to a single superordinate category that encompasses both animal-bats and baseball-bats. The current study explores the situations that lead children to postulate that a single word-form maps onto several distinct meanings, rather than a single superordinate meaning. Three experiments showed that adults and 5-year-old French children use information about the sampling of learning exemplars (and in particular the fact that they can be regrouped in two distinct clusters in conceptual space) to postulate homophony. This unexplored sensitivity and the very possibility of homophony are critically missing from current word learning accounts.

**Keywords:** word learning; homophony; concepts; constraints

---

\*Correspondence concerning this article should be addressed to Isabelle Dautriche, Laboratoire de Sciences Cognitives et Psycholinguistique, Ecole Normale Supérieure, 29 rue d’Ulm – P.J., 75005 Paris – France. E-mail: [isabelle.dautriche@gmail.com](mailto:isabelle.dautriche@gmail.com)

## Introduction

To learn a word, language learners must draw a link in their mental lexicon between a phonological form and its meaning. While many words conform with a one-to-one mapping between form and meaning, this is not always the case: a homophone is a phonological form associated arbitrarily with several meanings, each of which corresponds to a concept. For instance, the word form “bat” applies both to the concept ANIMAL BAT and to the concept BASEBALL BAT. Hence, homophones present children with a non-standard word learning situation, for which they need to discover that there is a decoupling between linguistic signals and concepts.

In order to examine what kind of challenge homophony brings into the word learning task, let us consider a typical word learning situation. Children do not observe associations between words and concepts; rather, they observe the co-occurrences of word forms and exemplars of their associated concepts. Thus, one major problem for the learner is to infer the meaning of the word from a set of exemplars that is consistent with an unbounded number of possible meanings (Quine, 1960). Existing theories of word learning have stressed the importance of prior knowledge to constrain the learning problem faced by the child (e.g., Bloom, 2001; Goodman, 1955; Markman, 1989). Such priors have been described at the level of concepts (what are the possible concepts our mind is ready to entertain) and at the level of word forms (what are the constraints on possible *form*-concept configurations).

All current accounts of word learning (associative learning accounts, e.g., Regier 2005; Yu & Smith 2007; hypothesis elimination accounts, e.g., Pinker 1989; Siskind 1996; Bayesian accounts, e.g., Frank et al. 2009; Piantadosi et al. 2012; Xu & Tenenbaum 2007) assume that learners rest on two main assumptions: First, the structure of the language is clear, that is, involves transparent, uniform mappings between forms and concepts, leading to one-to-one correspondence across these domains (a constraint at the level of word forms). Second, those concepts are *convex* (a constraint at the level of concepts), that is, whose members form a group that share a common set of properties that holds them to be contiguous in conceptual space (see further the notion of concept convexity in Gärdenfors, 2004; and the related notion of conceptual coherence in Murphy & Medin, 1985). For instance, a concept such as CAR OR WATER, is not a proper candidate for a concept because its members would be drawn from two disjoint sets which do not form a convex cluster of entities in conceptual space. Yet there is a sense in which this convexity constraint does not translate at the level of word forms: if a phonological form is used to refer to two objects A and B, it does not imply that all objects between A and B in conceptual space can also be labeled by the same form, as in the case of homophones (see discussion in Dautriche & Chemla, *submitted*). In other words, the extension of the meaning(s) of a word form is not necessarily convex.

Evidence for a convexity constraint at the level of concepts comes from several experimental studies (Dautriche & Chemla, *submitted*; Xu & Tenenbaum, 2007). In these studies, the conceptual space is defined over a tree-structured representation of entities by clustering a set of entities based on their similarity. Subtrees correspond to categories that words could label at different levels of granularity (e.g., cat, feline, mammal, animal). When exposed to a set of learning exemplars uniformly sampled from a category, adults extend the label to the minimal subtree including all the exemplars (Xu & Tenenbaum, 2007). For example, when presented with three “feps” labeling three Dalmatians, adults readily extend “fep” to the set of all Dalmatians, would they be presented with a Dalmatian, a Labrador and a German-shepherd they would extend the label to the set of all dogs. Yet, when presented with exemplars clustered at two distant positions in conceptual space, such as {two primates, one mushroom}, adults did not extend the label to

all objects falling within a convex category encompassing all exemplars (i.e. LIVING BEINGS), rather, they preferred to restrict the label to members of two disjoint subcategories (i.e. PRIMATE and MUSHROOM). Furthermore, this effect is strengthened by the manipulation of specific linguistic evidence that signals or bans homophony, such that the occurrence in some particular constructions (so called *zeugmas*, Zwicky & Sadock 1975). The interpretation is the following. When encountering novel words, adults are first guided by the idea of a one-to-one mapping between words and concepts, and convexity of concepts thus yields word forms with convex extensions. Yet, when observing evidence against the convexity of the set of entities that may be labelled by a given word form (i.e. the presence of two distinct convex clusters of learning exemplars in conceptual space), adults postulate homophony, that is they prioritize concept convexity over the possibility that a word form maps onto a single concept (Dautriche & Chemla, *submitted*).

An important question is whether concept convexity influences word learning during language development. Do children also expect word forms to refer to convex concepts? In case this seems untenable, do they also know about the backup strategy, homophony? Much developmental work suggests that children start with the assumption that one word form maps onto exactly one concept (Slobin, 1973, 1975) and that they follow some convexity constraint in that they expect concepts to group objects that share a common property. For instance, once children have mapped one form to an object, they will extend it to other objects that share the same ontological kind (*the taxonomic constraint*, Markman & Hutchinson, 1984), or the same shape (*the shape bias*, Landau et al., 1988). Yet, it remains open to question how children react to cases where concept convexity is challenged (see however Plunkett et al., 2008), i.e. when there is evidence that the extension of a word form is not convex: do they postulate homophony to maintain concept convexity (an important assumption to form concepts in the first place), do they prioritize sticking to a one-to-one mapping assumption between word form and meaning (an important assumption to learn words), or do they simply ignore the presence of two smaller convex clusters of exemplars and save their prior assumptions about both word forms and concepts by simply postulating a broad meaning for the word in question?

Relevantly, it has been experimentally demonstrated that 3- to 9-year-old children have difficulty in learning homophones (Casenhiser, 2005; Doherty, 2004; Mazzocco, 1997). In particular they find it more difficult to learn a second meaning for a word they know (thus a homophone, e.g., learning that the familiar word form "door" also labels an unfamiliar object) than to learn a completely novel word (e.g., learning that "blick" labels an unfamiliar object). This suggests that children are slower to learn secondary meanings of homophones than to learn novel words, consistent with the idea that children prefer to preserve a one-to-one mapping between forms and meanings. Yet another possibility is that homophone learning is difficult in these conditions because the homophones are chosen such that one meaning is already available to children and thus interferes with their ability to learn a novel meaning for the same word form. Indeed even when both meanings of a pair of homophones are known to children before the experiment, they find it difficult to retrieve the less frequent meaning of the pair when the most frequent meaning is activated, but have less trouble to do so when the task provides greater contextual support for the less frequent interpretation (Beveridge & Marsh, 1991; Campbell & Bowe, 1977; Rabagliati et al., 2013). That is, in experiments where a second meaning novel meaning is taught for a known word form, the competition between the known first meaning and the novel secondary meaning may mask children's ability to consider homophones as a possible option.

The present study addresses exactly this key problem by testing the simultaneous acquisition of two meanings for a single word form. When learning a homophone, such as "bat", learners will observe several exemplars of animal-bats and several exemplars of baseball-bats, all linked to the same word form "bat".

In such a case, if learners hypothesize that this word form applies to a single, convex concept, as most word forms do, they would never discover homophony. Instead, they could consistently postulate that “bat” refers to some superordinate, coherent category encompassing both animal bats and baseball bats, just like a word like “thing” does. However, if “bat” was indeed linked to such a broad category, it is likely that learners would have observed many things that are called “bat” but are neither animal bats nor baseball bats (a *uniform* distribution of exemplars drawn from the superordinate category of “things”) rather than having observed only animal bats and baseball bats (a *bimodal* distribution of exemplars within the superordinate category).

We thus ask whether children capitalize on the *sampling distribution of the learning exemplars* to postulate homophony. To our knowledge, this is the first study that looks at the acquisition of multiple meanings for a new word by children. In Experiment 1, we combined two tests (inspired from Srinivasan & Snedeker, 2011 and Xu & Tenenbaum, 2007) to replicate previous results with adult participants, circumscribing the situations in which homophony emerges. In Experiment 2, we exported our experimental procedure with children, showing that they also refrained from associating a label to a broad set of entities encompassing all learning exemplars when they form two distinct convex clusters. Experiment 3 provides a control with adults to discard the possibility of a superficial explanation for part of our effect. Altogether, our results suggest that children by the age of 5 use information about the sampling distribution of learning exemplars to discover whether a novel word form is associated with one or several meanings. Just like adults, children expect that words, but not word forms, refer to convex concepts and form lexical representations that follow this constraint, in essence showing early awareness that homophony is a possibility in natural languages.

## Experiment 1

The experiment consisted of two testing phases: the *extension test* and the *representation test*. The extension test was similar to Xu & Tenenbaum (2007) and Dautriche & Chemla (*submitted*): participants were taught novel labels from an alien language (e.g., “blicket”) for animal categories and were asked to extend this label to test items. We manipulated whether the set of exemplars they observed formed either a uniform or a bimodal distribution of the minimal superordinate category encompassing all the exemplars. If participants are sensitive to this sampling information, we predicted that they should be less likely to extend the label to all objects that are in the superordinate category when the exemplars form a bimodal distribution compared to when they form a uniform distribution.

During the representation test, we tested whether participants represented the meaning(s) of the word they have just been taught as two separate lexical entries (i.e. homophony) or as a single lexical entry. The procedure was similar to Srinivasan & Snedeker (2011): participants were taught that a subset of the examples previously shown were wrongly labelled and are in fact labelled by *another* word in that language (*the corrected label*, e.g., “these are not blickets, these are feps”). When the exemplars formed a bimodal distribution (as in the case of homophones, e.g., two animal-bats and two baseball-bats), the corrected label corresponded to one of the meanings of the initial word (e.g., “fep” labelled the two animal-bats). When the exemplars formed a uniform distribution (e.g., one animal-bat, one tree, one car, one-baseball bat), the corrected label applied to two of the exemplars (e.g., “fep” labelled one animal-bat and one car). Participants were then tested on their extension of the corrected label. If a bimodal distribution of exemplars is sufficient to trigger homophony, and assuming that participants can rely on the independence of the two meanings of a homophone, they should restrict the corrected label to the subcategory for which they have

evidence (e.g., “fep” refers to animal-bats) and not extend it to the broader category (e.g., exclude baseball-bats). On the opposite, if participants readily extend the corrected label to the unattested meaning (or at least as much as in the uniform condition), this would suggest that they interpreted the initial word as referring to a broad category, and not as a homophone.

## Method

### Participants

Nineteen adults were recruited from Amazon Mechanical Turk (6 Females;  $M = 37$  years; all native speakers of English) and were compensated \$0.4 for their participation. One additional participant was excluded because he did not provide any answer.

### Procedure and display

Adults were tested online. They saw the pictures of two aliens, one blue and one red, both coming from the same planet. They were instructed that they would be exposed to words from their language and would have to select images that correspond to those words. In the instructions, they saw an example of a trial with the pictures and the label used for the training trial.

The trials followed the time course schematically represented in Figure 1. In the learning phase, 4 learning exemplars were displayed as a combination of a picture and a prompt underneath each of them (e.g., “This is a blicket”), allegedly pronounced by the blue alien who was pictured at the bottom of the screen.

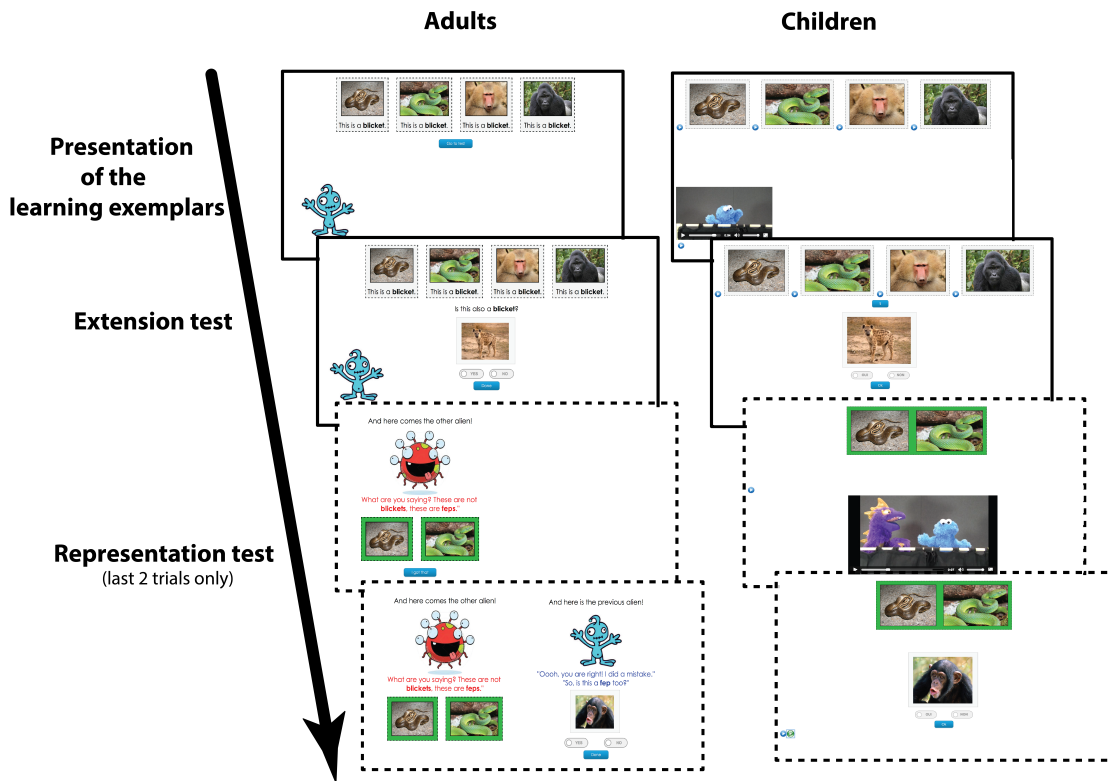
The *extension test* started as soon as adults pressed a “Go to test” button placed below the exemplars. Participants were presented with 12 test pictures displayed one-by-one below the 4 learning exemplars and asked whether this test item could be labelled by the novel word (e.g., “Is this also a blicket?”). Participants could answer by clicking on a “yes” or “no” button on the screen. When the response was “yes”, the picture frame became green, if it was “no”, the picture frame became red. Once the response was validated by participants by pressing a “Done” button, the test continued to the next test picture.

In the last two test trials, the extension test was followed by a *representation test*. On the left side of the screen, participants saw 2 of the 4 learning exemplars, highlighted in a green frame, with the red alien appearing with the prompt “What are you saying these are not blickets, these are feps!”.<sup>1</sup> Once participants pressed the “I got that” button, the blue alien re-appeared on the right side of the screen recognizing his mistake and asking whether the corrected label could apply to 3 novel test pictures presented one-by-one (e.g., “Ooooh you are right! I made a mistake, Is this a fep too?”). Participants validated their answer by pressing a “Done” button before moving to the next test picture.

At the end of the experiment, there was a final questionnaire asking participants about their age, native language and country.

---

<sup>1</sup>We chose to rename two of the learning exemplars instead of one following a pilot study conducted with adults. When a single exemplar was re-labelled, adults were confused whether the corrected label applied only to this specific instance or to the whole category of the initial label.

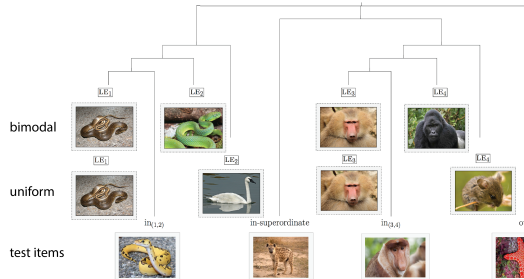


**Figure 1: Time course of a trial in experiments 1 and 2.** Adults were tested online while children were tested in their preschool. The mode of presentation differed between the two age groups: adults learned words through written prompts while children saw videos. The time course was the same for both groups. Participants first saw the 4 learning exemplars for the novel word presented by one of the aliens. In the extension test: participants then saw 12 test pictures presented one-by-one and were asked whether the image corresponds to the word just learned. In the last two trials, the extension test was followed by a representation test: first participants saw two of the learning exemplars being renamed with another label (the corrected label) by a second alien. Then participants see another set of 3 test pictures presented one-by-one and are asked whether each of them can also be named with the corrected label.

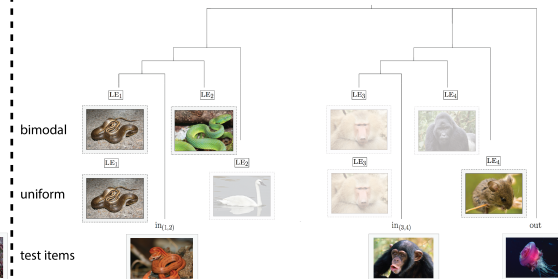
## Conditions

Each participant saw 1 training trial and 4 test trials: 2 *uniform* and 2 *bimodal* trials. For a simple and schematic explanation, we refer the reader to Figure 2 which represents the structure of test trials and introduces visually the associated terminology.

## 1. Extension test



## 2. Representation test



**Figure 2: Schema of the structure of a trial in conceptual space for the extension test (1) and the representation test (2).** The first row of pictures corresponds to the configuration of the learning exemplars ( $LE_1$ ,  $LE_2$ ,  $LE_3$ ,  $LE_4$ ) in the bimodal condition and the second row to the configuration of the learning exemplars in the uniform condition. The third row corresponds to the test items.

**Training trial.** The training trial was the same for all participants: the 4 learning exemplars were the pictures of 4 animals (a dog, a goat, a pig and a cow) and the 4 test items were two animals (a cat, a horse) and two plants (a tree, a pumpkin). It was designed so that participants understand readily the task, inviting them to extend the novel label to the two animals but not to the two plants.

**Test trial.** The key factor differentiating the two test conditions (*uniform* and *bimodal*) concerns the distribution of the learning exemplars ( $LE_1$ ,  $LE_2$ ,  $LE_3$ ,  $LE_4$ ) in conceptual space (here a tree-structure).

- In the *uniform trials*, the learning exemplars formed a uniform distribution sampled from a superordinate category such that all learning exemplars are about the same distance from one another.
- In the *bimodal trials*, the learning exemplars formed a bimodal distribution sampled from two independent subcategories belonging to the superordinate category such that they formed two clusters of exemplars: ( $LE_1$ ,  $LE_2$ ) and ( $LE_3$ ,  $LE_4$ ).

During the extension test (Figure 2.1), the 12 test items were either:

- *out*: out of the superordinate category formed by the 4 exemplars (4 items).
- *in*: in one of the two subcategories (2 items in between  $LE_1$  and  $LE_2$  and 2 in between  $LE_3$  and  $LE_4$ )
- *in-superordinate*: in the superordinate category but not in any subcategory (4 items)

During the representation test (Figure 2.2), another label (the corrected label) applied to 2 of the learning exemplars:  $LE_1$  and  $LE_4$  in the uniform trials and  $LE_1$  and  $LE_2$  in the bimodal trials. The 3 test items were either:

- *out*: out of the superordinate category formed by the 4 exemplars of the initial word.



- $in_{(1,2)}$ : in the subcategory formed by  $LE_1$  and  $LE_2$ .
- $in_{(3,4)}$ : in the subcategory formed by  $LE_3$  and  $LE_4$ .

## Materials

Our stimuli relied on a set of to-be-learned labels and taxonomically organized objects.

**Labels.** We chose 7 phonotactically legal non-words of English that were not repeated across trials: 5 for the initial labels *blicket*, *smirk*, *zorg*, *moop*, *tupa* and 2 for the corrected labels *kaki*, *fep*.

**Objects in conceptual space.** Participants were tested on a set of 100 animals organized into a taxonomic hierarchy extracted from NCBI (<http://www.ncbi.nlm.nih.gov>) to obtain an objective measure of similarity between the different items as in (Dautriche & Chemla, *submitted*).<sup>2</sup> For each item, we selected 3 color photographs showing the animal in its natural background.

## Presentation and trial generation.

The order of the trials as well as the pairing between the labels and the set of learning exemplars was fully randomized and differed for each participant. We created 2 lists of trials, such that each uniform trial had a corresponding bimodal trial in the other list. That is  $LE_1$ ,  $LE_3$  and the test items were common between a pair made of a uniform and a bimodal trial, while  $LE_2$ ,  $LE_4$  varied to make the trial uniform or bimodal. All trials were generated automatically following the algorithmic constraints described in the supplemental material and selected following pilot data on adults (the full list of trials is available in the supplemental material). Participants were randomly assigned to one of the two lists of trials.

## Data analysis

Analyses were conducted using the `lme4` package (Bates et al., 2014) of R (R Core Team, 2013). In a mixed logit regression (Jaeger, 2008), we modeled the selection of a test item (coded as 0 or 1) independently for the extension test and the representation test. The extension test model included two categorical predictors with their interaction: Test Item (*out*, *in*, *in-superordinate*) and Sampling Condition (uniform vs. bimodal) as well as a random intercept and random slopes for both Test Item and Sampling Condition and their interaction for participants and trial pairs.

The representation test model included as well two categorical predictors with their interaction: Test Item (*out*,  $in_{(1,2)}$ ,  $in_{(3,4)}$ ) and Sampling Condition (uniform vs. bimodal) with a random intercept and random slopes for Sampling Condition for participants.<sup>3</sup> As can be seen in Figure 4, participants were at ceiling in selecting  $in_{(1,2)}$  in the bimodal condition. This impacted the log-estimation behind the logit model for the representation test, such that the estimates and the standard errors calculated by the model

<sup>2</sup>Species for which there is a clear difference between scientific and subjective popular taxonomy were excluded (e.g., sea mammals: counter-intuitively, dolphins are closer to elephants than to sharks).

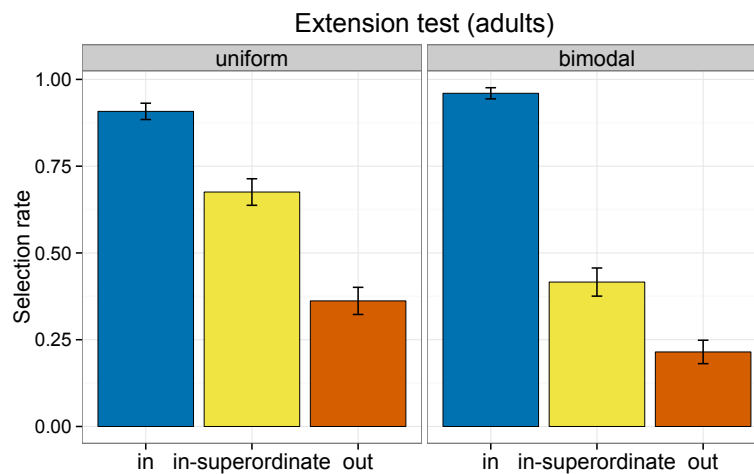
<sup>3</sup>There was no random slopes for Test Items because it leads to implausible estimates and standard error even after correction (see thereafter). There was also no random effect for trial pairs since the two last trials could be different for each participant.

were implausibly big. To get rid of these ceiling effects in a highly conservative way we introduced random noise: we run the same analysis on a modified dataset where we changed randomly 10% of the responses given for  $in_{(1,2)}$  in the bimodal condition (2 responses).

## Results

### Extension test

Figure 3 reports the average proportion of selection of each test item by sampling condition (uniform vs. bimodal) during the extension test.



**Figure 3:** Proportion of choice of each test item during the extension test averaged for each trial condition (uniform vs. bimodal) for adults in Experiment 1. Error bars indicate standard errors of the mean.

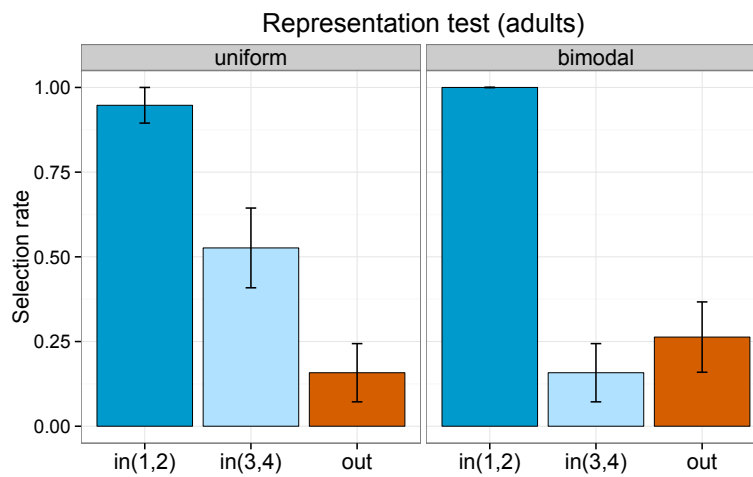
Participants were sensitive to the distribution of the learning exemplars: they selected more *in-superordinate* items in the uniform condition than in the bimodal condition ( $M_{uniform} = 0.68$ ,  $SE = 0.04$ ;  $M_{bimodal} = 0.45$ ,  $SE = 0.04$ ;  $\beta = -3.38$ ,  $z = -2.48$ ,  $p < .05$ ). We note that the distribution of the learning exemplars also affected the selection rate of *out* items: they selected more *out* items in the uniform compared to the bimodal condition ( $\beta = 1.79$ ,  $z = 2.03$ ,  $p < .05$ ). Yet, this is expected following the size principle documented by Xu and Tenenbaum (2007): the boundaries of the superordinate category defined by the 4 learning exemplars in the uniform condition are less sharp (for an equal number of exemplars) than the boundaries of the subcategories in the bimodal condition. As a result, there is more uncertainty about the correct level of generalization in the tree-structured hierarchy leading to a slightly higher selection rate of *out* items in the uniform condition.

However the sampling distribution of the exemplars modulated participants' responses beyond the size principle effect since we observe two interaction effects: the difference between the selection rate of *in-superordinate* items and *in* items was greater in the bimodal than in the uniform condition ( $\beta = 5.70$ ,  $z = 3.23$ ,  $p < .01$ ); similarly the difference between the selection rate of *in-superordinate* items and *out* items

was marginally smaller in the bimodal than in the uniform condition ( $\beta = -1.58, z = -1.67, p < .1$ ). This suggests that participants were more inclined to extend the label to all objects in the superordinate category including all the exemplars when the exemplars formed a uniform distribution than when they formed a bimodal distribution.

## Representation test

Figure 4 reports the average proportion of selection of each test item by trial condition (uniform vs. bimodal) during the representation test.



**Figure 4:** Proportion of choice of each test item during the representation test averaged for each trial condition (uniform vs. bimodal) for adults in Experiment 1. Error bars indicate standard errors of the mean.

Crucially, the distribution of the exemplars influenced participants' representation of the initial word: in the bimodal condition, participants were less likely to extend the corrected label to unattested items ( $in_{(3,4)}$ ) than in the uniform condition ( $M_{uniform} = 0.53, SE = 0.12; M_{bimodal} = 0.16, SE = 0.08; \beta = -5.05, z = -3, p < .01$ ). This is compatible with a homophonous representation of the initial word in the bimodal condition, where only one of the two meanings ( $in_{(3,4)}$ ) has been affected by the later correction.

## Discussion

The sampling distribution of the exemplars modulated adults' interpretation of a novel word: when the exemplars formed a uniform distribution, participants were more likely to extend its label to all objects falling in the superordinate category containing all the exemplars than when the exemplars formed a bimodal distribution (extension test).

Yet, one may worry that participants did not form a lexical representation but rather extended the label based on similarity to the learning exemplars: they selected more *in-superordinate* items in the uniform than in the bimodal condition simply because, on average, *in-superordinate* items may be closer to the

learning exemplars in the uniform than in the bimodal condition. Yet that's not the case: recall that in our similarity space (Figure 2), the common ancestor of *in-superordinate* items with any pair of learning exemplars, i.e. (LE<sub>1</sub>, LE<sub>2</sub>) or (LE<sub>3</sub>, LE<sub>4</sub>), is the same in the uniform and the bimodal condition. Certainly participants' responses were in part guided by similarity of the test items to the learning exemplars: in the bimodal condition during the extension test, adults selected *in-superordinate* items at a higher rate than *out* items ( $\beta = -2.30, z = -1.99, p < .05$ ). Yet it is sufficient for our purpose to note that this does not account for the entirety of our effect.<sup>4</sup>

An interesting question is whether observing the exemplars in a bimodal distribution was sufficient for participants to form homophonous form-meaning representations. Indeed, there may be two interpretations of the results of the extension test:

1. Participants formed words' representations that respect concept convexity. In the bimodal condition, participants postulated homophony: they associated the novel word with two independent meanings, each corresponding to a convex concept (e.g., PRIMATE and SNAKE).
2. Participants accepted that a word's meaning could be a set of disconnected concepts: in the bimodal condition they associated the novel word to a single, disjoint concept (e.g., PRIMATE OR SNAKE).

The results of the representation test favor the first possibility. When presented with a bimodal distribution of exemplars (e.g., 2 primates and 2 snakes) labelled by a single word "blicket", participants interpreted the corrected label based on its taught meaning alone (e.g., the 2 snakes but not the 2 primates) suggesting that they preferentially understood "blicket" as a word form associated with two homophonic words, rather than as a single word with a single discontinuous meaning.

All in all, we replicate previous results (Dautriche & Chemla, *submitted*) showing that the distribution of learning exemplars interacts with constraints on concept convexity to form different form-meaning representations. When the exemplars form a uniform distribution, participants are more likely to associate the word to a single convex meaning that encompasses all the learning exemplars (and every entity in between them). Yet, when the exemplars form a bimodal distribution, participants prefer to postulate homophony such that the novel word is associated to two convex meanings, rather than to a single, broad convex meaning or to a single discontinuous meaning. The critical question then, is, do children also postulate homophony when there is evidence that the exemplars of a word are distributed in two convex clusters in conceptual space? Experiment 2 investigated this question by adapting the design of Experiment 1 to French preschoolers.

---

<sup>4</sup>Note that similarity to the exemplars also played a role in the representation test, in which participants selected more  $in_{(1,2)}$  items than  $in_{(3,4)}$  items in the uniform condition ( $\beta = -4.01, z = -2.61, p < .01$ ) simply because in the tree structure for uniform trials (Figure 2),  $in_{(1,2)}$  is closer to LE<sub>1</sub> than  $in_{(3,4)}$  to LE<sub>4</sub>, thus more likely to be selected as an instance of the corrected label.

## Experiment 2

### Method

#### Participants

Twenty-one 5-year-old monolingual French speaking children (5;1 to 6;1,  $M_{age} = 5;6$ , 10 girls) were tested in a public preschool in Paris. Their parents signed an informed consent form. Three additional children were tested but not included in the analysis because they systematically responded *yes* ( $n = 1$ ) or *no* ( $n = 2$ ) without even looking at the test pictures or responding before they appeared on the screen.

#### Procedure and display

The experiment was identical to Experiment 1 except that we used videos instead of written prompts with children (see Figure 1).

Children were tested individually in a quiet room in their preschool. During the experiment, children sat next to the experimenter, in front of a computer and wore headphones to listen to the stimuli. Before the experiment began, children watched a video where two alien puppets introduced them with the task. The two puppets presented themselves as Boba and Zap, and told the children that they were coming from another planet where they speak a different language, so they would teach the children words of their language. Once a child demonstrated to the experimenter that (s)he understood the task, the experiment started. In each trial, children saw 4 learning exemplars, presented one-by-one as the combination of a picture and a video of Boba labeling the picture with a non-word “Ça, on appelle ça une bamoule!” *This, we call it a bamoule*. Each learning exemplar was displayed on the screen when the experimenter clicked on a button. Once children saw all 4 learning exemplars, they saw a last video where Boba asked them to repeat the word. This was to ensure that children were on task and for the experimenter to know which word was used by the puppet (as the words were randomly assigned to a set of learning exemplars and the experimenter could not hear the stimuli).

During the *extension test*, children were presented with the 12 test pictures displayed one-by-one below the 4 learning exemplars. For each of them, the experimenter asked: “Est ce que tu penses que ça s’appelle une bamoule?” *Do you think it is called a bamoule?*. When the child answered, the experimenter clicked accordingly on the “yes” or “no” button on the screen. Children could change their mind if they wanted within a few seconds after their answer or longer if the experimenter saw that they were still hesitating. Once a response was validated by the experimenter, the test continued to the next test picture.

In the last two test trials, the extension test was followed by a *representation test*. Children saw 2 of the 4 learning exemplars grouped in a frame together with a video where the second puppet, Zap, scolded the first one, Boba, for using the wrong word for the two exemplars displayed “C’est pas des bamoules ça! C’est des torbas!” *These are not bamoules! These are torbas* (the whole script for this video can be found in the supplemental material). Boba then, acknowledged his mistake. During the dialogue, the frame of the pictures was blinking in green. At the end of the video, the experimenter asked the child what happened and replayed the video if the child did not understand the video or did not remember the novel word. Children were then tested whether the corrected label could apply to 3 novel test pictures. Each of the test picture

was displayed below the two learning exemplars and the child was asked by the experimenter “Est ce que tu penses que ça, ça s’appelle un torba?” *Do you think that this is called a torba?*

At the end of the experiment, there was a final video where the two puppets said good-bye to the child. The whole Experiment lasted about 15 min. All sessions were audiotaped.

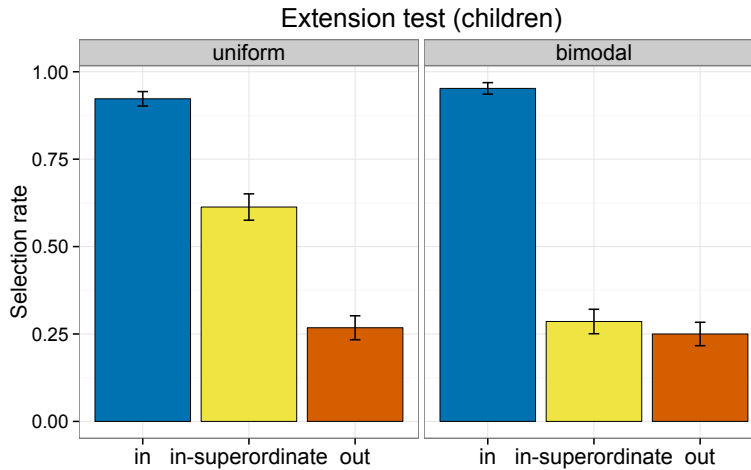
## **Conditions, Materials, Presentation and trial generation, Data analysis.**

Everything was similar to Experiment 1, except that the experiment was in French and hence the set of non-words consisted of 9 phonotactically legal non-words of French: *bamoule* was always used in the training trial. From the remaining 8 non-words that were used in the test trials, half of them were bisyllabic and were used as the initial label (*toupa, fimo, lagui, yoshi*), the other half were trisyllabic and were used as the corrected labels (*midori, cramoucho, didolu, baboocha*). The difference in the number of syllables was introduced to make it easier for children to distinguish between the initial and the corrected labels.

## **Results**

### **Extension test**

As shown in Figure 5, children and adults behaved in the same way: children selected more *in-superordinate* items in the uniform condition than in the bimodal condition ( $M_{uniform} = 0.61$ ,  $SE = 0.04$ ;  $M_{bimodal} = 0.28$ ,  $SE = 0.04$ ;  $\beta = -2.56$ ,  $z = -3.73$ ,  $p < .001$ ). The distribution of the learning exemplars did not affect any other test items for children (all  $ps > 0.7$ ) resulting in two interaction effects: the difference between the selection rate of *in-superordinate* items and *in* items, was greater in the bimodal than in the uniform condition ( $\beta = 2.79$ ,  $z = 2.30$ ,  $p < .05$ ); similarly the difference between the selection rate of *in-superordinate* items and *out* items, was smaller in the bimodal than in the uniform condition ( $\beta = -3.28$ ,  $z = -4$ ,  $p < .001$ ). This suggests that children were more likely to extend the label to all objects in the minimal subtree containing all the exemplars in the uniform condition compared to the bimodal condition.



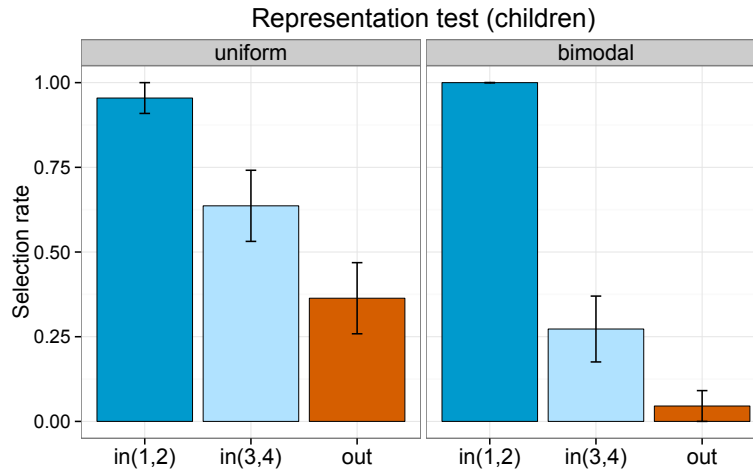
**Figure 5:** Proportion of choice of each test item during the extension test averaged for each trial condition (uniform vs. bimodal) for children. Error bars indicate standard errors of the mean.

For uniform trials, we replicate previous results (Xu & Tenenbaum, 2007): children were more likely to extend the label to all objects in the minimal subtree containing all the exemplars than to objects that are out of the subtree (*in* vs. *out*,  $\beta = -5.39, z = -6.28, p < .001$ ; *in-superordinate* vs. *out*,  $\beta = -3.01, z = -3.82, p < .001$ ). We note that the distance of the test items to the learning exemplars affected children’s responses: they selected more *in* items than *in-superordinate* items ( $\beta = -2.29, z = -2.80, p < .01$ ) suggesting that children’s extension of the label may be in part guided by similarity of the test items to the learning exemplars.

For bimodal trials, children were more likely to extend the label to objects that were in one of the two subcategories (*in*) than to other objects that were either in the superordinate category containing all the exemplars but out of the subcategories (*in-superordinate*,  $\beta = -5.12, z = -4.61, p < .001$ ) or out of the superordinate category (*out*,  $\beta = -4.85, z = -8.01, p < .001$ ). There was no difference between the selection rate of *in-superordinate* items and *out* items for children ( $p > 0.7$ ) suggesting that children excluded *in-superordinate* items from the extension of the word.

## Representation test

The distribution of the exemplars influenced children’s representation of the initial word (see Figure 6): in the bimodal condition, children were less likely to extend the corrected label to unattested items ( $in_{(3,4)}$ ) than in the uniform condition ( $M_{uniform} = 0.64, SE = 0.10$ ;  $M_{bimodal} = 0.28, SE = 0.10$ ;  $\beta = -1.66, z = -2.51, p = .01$ ).



**Figure 6:** Proportion of choice of each test item during the representation test averaged for each trial condition (uniform vs. bimodal). Error bars indicate standard errors of the mean.

While children and adults behaved statistically the same way ( $p > 0.2$ ), there were some visible differences. In particular, there was a main effect of Sampling Condition for children: children were more likely to select test items in the uniform than in the bimodal condition ( $\chi(1) = 8.20, p < .01$ ). They even selected more *out* items in the uniform than in the bimodal condition ( $\beta = 2.56, z = 2.20, p < .05$ ; no such difference was observed for adults,  $p > 0.2$ ). The selection rate of *out* items in the uniform condition is similar to their selection rate during the extension test. One may wonder why children selected it even less in the bimodal condition during the representation test. Children may be driven by a “size principle” (Xu & Tenenbaum, 2007): what determines whether the hypothesized extension of a label will have sharp boundaries is the number of consistent hypotheses with the exemplars. In the representation test, the category defined by the two exemplars is rather narrow (e.g., snakes) ensuring that very few meaning hypotheses are possible for the corrected word. During the extension test, there were more meaning hypotheses possible for the initial word: despite the fact that two exemplars were presented for each of the two subcategories (e.g., INSECTS and PRIMATES), the possibility that the initial word corresponded to the minimal superordinate category encompassing all the exemplars could still be entertained leading to a bigger uncertainty about the boundaries of the categories of the initial word compared to the boundaries of the corrected word.

## Discussion

These results suggest that children, just like adults, postulate homophony for a word when the learning exemplars formed a bimodal distribution, i.e. that the meaning of that word form is best represented as two independent convex clusters rather than a big cluster encompassing all the exemplars.

At this point, we would like to point out a weakness in the second of our tests, the representation test. Because the category including the re-labelled exemplars is wider in the uniform case (e.g., a snake and a mouse) compared to the bimodal case (e.g., two different kinds of snakes), this may be sufficient for children to extend the corrected label to more test items in the uniform compared to the bimodal condition: indeed, a chimp is more likely to be a “fep” when “fep” labels a snake and a mouse than when “fep” labels



two snakes. This could happen independently of what children had learned about the meaning of the initial word “blicket”. Thus, although our results reflect exactly what we could expect to observe if children had hypothesized that “blicket” was a homophone when observing a bimodal distribution of its exemplars, we cannot entirely rule out the possibility that children, and adults, responded to the representation test independently from the extension test, in which case the representation test would tell us nothing about what had been learned initially. To test whether participants responded to the representation test independently from the extension test, i.e., whether participants’ extension of the corrected label was independent from their representation of the initial word, we ran a control experiment with adults, in which we tested both participants’ representation of the initial word not only by testing their extension of the corrected label (“fep”) but also by testing their *updated* extension of the initial label (“blicket”).

## Experiment 3

In Experiment 3, adults learned a novel word, e.g., “blicket” (the initial label), from a set of exemplars, and were instructed that some of the exemplars were not “blickets” but “feps” (the corrected label).

When “blicket” applies to a bimodal distribution of exemplars (e.g., two animal-bats and two baseball-bats) and “fep” corresponds to one of the meanings of the initial word (e.g., the two animal-bats), if participants recruit what they have learned from “blicket”, they should restrict not only the corrected label to the subcategory for which they have evidence (e.g., “fep” refers to animal-bats and not to baseball-bats) as in Experiment 1, but they should also update their representation of “blicket” (e.g., “blicket” now refers to baseball-bats and not to animal-bats). On the opposite, if they readily extend “blicket” to animal-bat items, although these items are now labelled as “feps”, then this will suggest that their responses for “fep” and “blicket” are independent. When “blicket” applies to a uniform distribution of exemplars (e.g., one animal-bat, one tree, one car, one-baseball bat) and the corrected label applies to two of the exemplars (e.g., “fep” labelled one animal-bat and one car), if participants recruit their lexical representation of “blicket” during the representation test, they should, as in Experiment 1, be more willing to extend “fep” to all “blickets” (e.g., the broader category of animal-bat, tree, car and baseball bat) but crucially they should also update their lexical representation for “blicket” (e.g., excluding now the broader category of animal-bat, tree, car and baseball bat from it). On the contrary if participants’ lexical representation of “blicket” is unaltered, this would suggest that participants have independent lexical representations for “fep” and “blicket”.

In sum, for both conditions if participants’ extension of the corrected label, “fep”, is uninformed by what they learned about “blickets”, then their representation of “blicket” should be untouched. On the other hand, if participants recruited their lexical representation of “blicket” during the representation test, then they should update this lexical representation by excluding all the “feps” from it.

## Method

### Participants

Twenty adults were recruited from Amazon Mechanical Turk (8 Females;  $M = 33$  years; 19 native speakers of English) and were compensated \$0.4 for their participation.

## Procedure and display

The experiment was identical to the adult version in Experiment 1 except that during the representation test we tested their representation of the corrected word (as in Experiment 1) but also their representation of the initial word.

During the representation test, as before, on the left side of the screen the red alien labelled 2 of the 4 learning exemplars by another word “What are you saying these are not blickets, these are feps!”. On the right side of the screen, the blue alien seen during the extension text appeared with a test item and two prompts 1) “Ooooh you are right! I made a mistake, Is this a fep too?” (representation test for the corrected word) and below it 2) “Is this a blicket?” (representation test for the initial word). There was a “yes” or “no” button below each question to record participants’ answer.

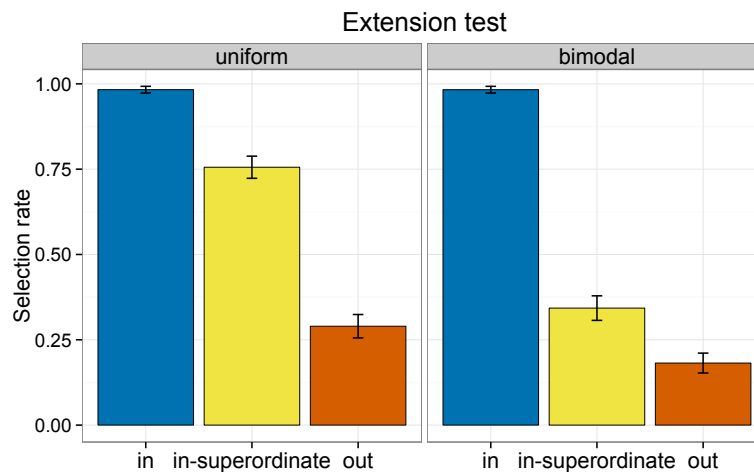
## Conditions, Materials, Presentation and trial generation, Data analysis.

Similar to Experiment 1.

## Results

### Extension test

Figure 7 reports the average proportion of selection of each test item by trial condition (uniform vs. bimodal) during the extension test.



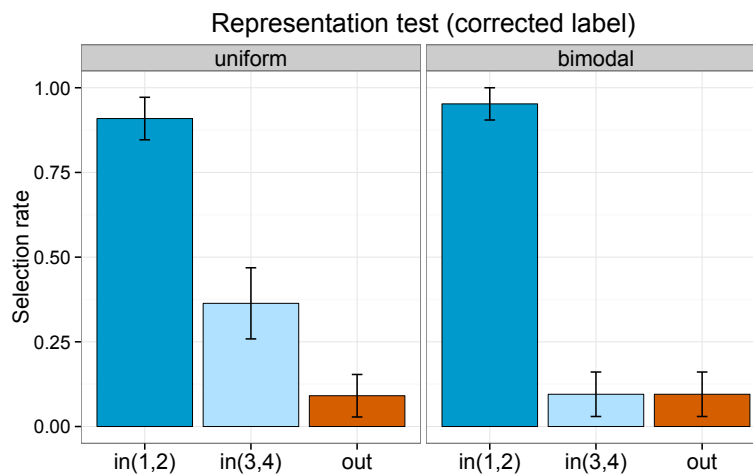
**Figure 7:** Proportion of choice of each test item during the extension test averaged for each trial condition (uniform vs. bimodal) in Experiment 2. Error bars indicate standard errors of the mean.

The extension test replicated the results of Experiment 1: participants’ responses were modulated by the distribution of the learning exemplars: participants chose more *in-superordinate* items in the uniform than in the bimodal condition ( $M_{uniform} = 0.75$ ,  $SE = 0.04$ ;  $M_{bimodal} = 0.34$ ,  $SE = 0.03$ ;  $\beta = -6.31$ ,  $z =$

$-3.57, p < .001$ ). The sampling distribution did not affect the choice of any other test items ( $ps > 0.1$ ). As a result, the difference between the selection rate of *in-superordinate* and *in* items was greater in the uniform than in the bimodal condition ( $\beta = 3.68, z = 1.98, p < .05$ ) and the difference between the selection rate of *in-superordinate* and *out* items was smaller in the bimodal condition compared to the uniform condition ( $\beta = -5.33, z = -2.51, p < .05$ ).

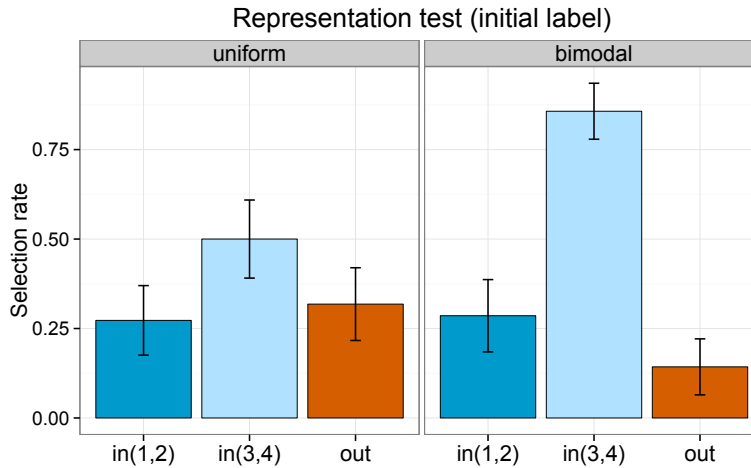
## Representation test

Figure 8 reports the average proportion of selection of each test item by trial condition (uniform vs. bimodal) during the representation test for the corrected label. This test also replicated the results of Experiment 1: in the bimodal condition, participants were less likely to extend the corrected label to unattested items ( $in_{(3,4)}$ ) than in the uniform condition ( $M_{uniform} = 0.36, SE = 0.10; M_{bimodal} = 0.1, SE = 0.06; \beta = -2.26, z = -2.19, p < .05$ ). The sampling distribution of the exemplars did not affect the responses to other test items ( $ps > 0.5$ )



**Figure 8:** Proportion of choice of each test item during the representation test averaged for each trial condition (uniform vs. bimodal) when participants were tested on the extension of the corrected label. Error bars indicate standard errors of the mean.

As shown in Figure 9, when tested on the initial label, participants selected more  $in_{(3,4)}$  in the bimodal than in the uniform condition ( $M_{uniform} = 0.5, SE = 0.11; M_{bimodal} = 0.86, SE = 0.08; \beta = 2.28, z = 2.70, p < .01$ ). This suggests that the sampling distribution of the exemplars affected participants' representation of both the corrected and the initial words. The sampling distribution did not affect the response of any other test items ( $ps > 0.1$ ).



**Figure 9:** Proportion of choice of each test item during the representation test averaged for each trial condition (uniform vs. bimodal) when participants were tested on the initial label. Error bars indicate standard errors of the mean.

In the uniform condition, participants refrained from applying the initial label to any of the test items that were within the superordinate category formed by the 4 exemplars ( $in_{(1,2)}$  and  $in_{(3,4)}$ ). This was evidenced by the fact that the selection rate of  $in_{(1,2)}$  and  $in_{(3,4)}$  as representatives of “blickets” was not different from their choice of the *out* items ( $p > 0.15$ ). This suggests that the corrected word overrode the representation of the initial word: all items previously labelled as “blickets” are now “feps”, thus “blicket” is not associated with any meaning.

In the bimodal condition, participants refrained from applying the initial label to test items that belonged to the subcategory of items that had been relabelled with the corrected label: they did not choose  $in_{(1,2)}$  as representatives of the initial label more often than *out* items ( $p > 0.2$ ); however, they readily extended the initial label to test items that were outside the subcategory of relabelled items ( $in_{(3,4)}$  vs. *out*:  $\beta = -4.52, z = -4.46, p < .001$ ). In other words, when exposed to two primates and two snakes all labeled “blicket”, if the two primates are relabelled “feps”, participants still consider that the new instances of snakes are valid instances of “blicket”.

## Discussion

When the initial word was corrected by another word (e.g., “these are not blickets, these are feps”), participants took into account what they have learned about the initial word (“blicket”) to comprehend the extension of the corrected word. We replicated the results from Experiment 1: in the uniform condition, participants were more likely to associate the corrected word to the superordinate category that spans all 4 learning exemplars than in the bimodal condition. In addition, participants also updated their representation of the initial word. Specifically, in the uniform condition, participants refrained to associate the initial label to all items that were previously labelled by it (as all are now in the extension of the corrected word, “fep”). Yet, when the exemplars formed a bimodal distribution, i.e. were taken from two clusters of exemplars  $C_1$  and  $C_2$  such that the corrected label applied only to one of the clusters of exemplars  $C_1$ , participants still considered test items belonging to  $C_2$  as valid instances of the initial label. This suggests that adults in

Experiment 1 recruited the representation of the initial word during the representation test.

However, the alternative explanation, i.e., that the representation of the corrected label “fep” is computed on the basis of the learning exemplars for “fep” only (independently of the representation of the initial label “blicket”), cannot be entirely ruled out. Participants could have extended the corrected label “fep” according to the distribution of its exemplars in the representation test (i.e., extend it to more test items in the uniform condition than in the bimodal condition simply because the category covered by the two exemplars in the representation test is wider in the uniform case) and then used this information to decide how “blicket” should be extended. Clearly if participants decide that a given item *I* is an instance of “fep”, they will be more willing to exclude it from the extension of “blicket” and this independently of how they decided that *I* is a “fep”. However, it should be noted that *out* items are excluded both from the extension of “fep” and the extension of “blicket”, suggesting that when responding to “blicket” in the representation test, participants are still influenced by their responses for “blicket” during the extension test (where they also excluded *out* items from the extension of the word). So this would suggest that, while participants may not recruit their initial representation of “blicket” to extend “fep” to test items in the representation test, they still recruit it (together with what they have learned about “fep”) to find the extension of “blicket” in the representation test. While we cannot entirely dismiss the possibility that participants would selectively attend to their initial representation of “blicket” in the representation test when responding to “blicket” but not to “fep”, our data provide little support for this hypothesis.

We conclude thus that the most likely interpretation of our results in Experiment 1 is that adults recruited the representation of the initial word during the representation test. Adults preferred, in the first phase, to postulate that the novel word carries homophony when it is learned from exemplars in a bimodal distribution. Given the similarity of the results between Experiments 1 and 2, one would be tempted to extend this interpretation to the children results from Experiment 2. At this point, however, a note of caution is necessary. One limitation of the present data is that we used our conclusion with adults to rule out a possible confound for Experiment 1 and 2 with adults and children. Yet we cannot exclude the possibility that this very explanation may still underly children’s response pattern. We leave it for future research to establish a more direct argument to explain both children’s and adults’ performance in these experiments.

## General Discussion

Children are sensitive to the sampling distribution of the learning exemplars when learning words (as in Xu & Tenenbaum, 2007). Yet, we demonstrate that this interacts with the kind of form-meaning representation children are ready to entertain. Observing a bimodal distribution of learning exemplars for a novel word indicated to our participants that the word was likely to have several meanings. Importantly, our results suggest that these meanings were stored *separately*, suggesting that children’s representation of the novel word in these conditions is very much similar to homophony (Srinivasan & Snedeker, 2011). This extends previous results from adults (Dautriche & Chemla, *submitted*) and suggests that when observing a bimodal distribution of exemplars for the same word form, children generate form-meaning representations, such as homophony, that respect concept convexity.

Current word learning accounts have documented and modeled paradigmatic cases of word acquisition, where a single form is associated with a single meaning. We pursued that enterprise by showing that less standard situations, such as homophony, can help highlight the key role of factors such as the sampling

distribution of exemplars. It also helps better understand the priors that may constrain and guide word acquisition, as we detail below.

## Missing factors in word learning accounts

Previous studies have shown that children are sensitive to sampling principles when learning words. Xu & Tenenbaum (2007) describe a *size principle*: children’s confidence in the boundary of the set of entities associated with a word increases as they observe more learning exemplars, even if they are all identical. Here we showed that children were sensitive to the distribution of learning exemplars in conceptual space, another statistical principle presumably following from the assumption that the exemplars of a word are sampled *randomly* from the underlying category (c.f Xu & Tenenbaum, 2007). Intuitively, in the case of homophones, we expect the label to occur with a set of exemplars  $S$  drawn from two distinct subcategories  $X_1$  and  $X_2$ . Thus,  $S$  would take the form of a *bimodal* distribution within the single superordinate category  $X$  encompassing the two subcategories (i.e.  $X$  is the minimal well-formed category such that  $X_1 \cup X_2 \subset X$ ). Yet if the word were to be associated with the whole  $X$ , we would expect the exemplars that are associated with the label (i.e.,  $S$ ) to be *uniformly* distributed within  $X$ . Our results suggest that children can use such sampling considerations to decide whether a word is associated with one category (standard case) or several categories (homophone), even when exposed to very few exemplars.<sup>5</sup>

There are other factors, not documented here, that may interact with the expectation that concepts are convex and could help children to identify that a word has several meanings. First, evidence for homophony may come from other words in the lexicon. Adults are less likely to extend a label (e.g., “blicket”) to an entity, even if this entity falls right between the learning exemplars for the label, when this entity also falls close to some entity labelled by another word (e.g., “fep”) (Dautriche & Chemla, *submitted*). Intuitively, the interfering label provides further evidence for the presence of two distinct clusters of exemplars in conceptual space, that are separated by another concept labelled by another word (“fep” in the example). This suggests that learners have expectations not only about how words occupy the conceptual space, but also about how they *share* the conceptual space. Similarly, children assume that word extensions are mutually exclusive (Markman & Wachtel, 1988), and may thus possibly use the presence of other words in their lexicon together with a constraint on concept convexity to discover that a word is likely to have several meanings.

Second, some linguistic constructions may be helpful to discover homophony (or the absence of homophony). For instance, children could notice that words mapping to a single meaning commonly appear in some plural sentences where homophones never appear (e.g., “These are two bats” pointing at one baseball-bat and one animal-bat). And this is so for reasons one can understand: a single phonological form cannot be used to refer to two words at the same time, even if the two words are homophonic.<sup>6</sup> Adults have been shown to use such constructions to assess homophony (Dautriche & Chemla, *submitted*), and children may also be sensitive to such linguistic evidence.

These factors may help learners to identify words with multiple meanings. Yet, they also raise immediate challenges for current word learning accounts. For instance, we assumed until now that children understood which object is referred to by the word in context. Yet, in the real world, the label is uttered

---

<sup>5</sup>One may argue that the low number of exemplars was not a limitation in our task because it could be compensated by pragmatic considerations: participants may expect that the learning exemplars were not drawn at random by the aliens, but rather that the aliens in our stories were trying to be informative about the meaning of the words and chose their exemplars optimally.

<sup>6</sup>Note that such a sentence is not agrammatical but *zeugmatic* (Zwicky & Sadock, 1975).

in a complex visual environment where the true referent is likely to be confounded with other possible referents present at the same time (the *mapping problem*, Quine, 1960). Thus it is likely that the set of exemplars for a label contains *outliers*, i.e. items that are outside of the true extension of the word, because the child would have failed to narrow down the true referent of the word. As a result, the set of exemplars would certainly form a multimodal distribution (e.g., a set of banana exemplars along with a dog exemplar – that happened to eat a banana during one of the learning event). The challenge for the learner is thus to distinguish between outliers of the true meaning of the word and representative examples of a new word meaning. General principles such as the convexity of concepts and one-to-one mapping between word forms and concepts may help discard noise of this type. However, if these general principles allow exceptions, as our study of homophony reveals, they may hardly help disentangle signal (of homophony) from noise.

## Missing priors in word learning accounts

Our results suggest that children expect meanings to be convex, and are willing to postulate homophony rather than breaking this constraint (postulating a disjoint meaning) or than enforcing that convexity constraint at all cost (postulating a broad lexical entry for problematic words).

This contradicts current word learning accounts which, technically, transpose the notion of convexity from the level of concepts to the level of word forms, assuming that word forms link to concepts in a one-to-one fashion. Accordingly, none of the current accounts allow for the possibility that children can associate word forms with multiple meanings. As a matter of fact, many developmental studies have documented that preschoolers have notable difficulties in learning homophones (Casenhiser, 2005; Doherty, 2004; Mazzocco, 1997). In these studies, the encounter of the second meaning of a homophone is simulated by using familiar words (e.g., “snake”) to refer to novel referents (e.g., an unfamiliar object). Yet, we suggest that children’s failure in these studies does not reflect an excessive reliance on a one-to-one mapping between form and meaning, but rather insufficient executive skills for such a task; as the current results show, children have no problem learning homophones when they have to learn the two meanings simultaneously. Learning homophones in our study may be easier than learning a second meaning for a known word because children do not have to inhibit a highly active word representation for one of the meanings (Khanna & Boland 2010, see also Choi & Trueswell 2010; Novick, Trueswell, & Thompson-Schill 2010). This suggests that children’s difficulty in learning homophones may have been previously overrated and that endowing the learning system with a strict one-to-one form-meaning mapping constraint cannot solely account for the mechanism underlying the acquisition of homophones. Certainly it may still be possible that a one-to-one form-meaning mapping constraint is guiding word learning at earlier stages of language development, as 5-year-olds may already have learned to relax this constraint to accommodate more challenging form-meaning mappings, such as homophony. Yet, if this is the case, current word learning accounts should be able to explain how children depart from this default assumption.

The present work thus has important implications for current word learning accounts. When a label seems to apply to a disjoint set of objects, the learner has two options: 1) postulate homophony (i.e. the possibility that a word form maps onto two distinct meanings) or 2) follow a one-to-one form-meaning mapping constraint and postulate that the label is a single word that applies to a larger set of objects (the category that covers all the positive instances of the label). Most accounts predict that 2) is the default, but this has to be refined since children eventually learn homophones (e.g., it is likely that English preschoolers

know both meanings of the word form “bat”; see also de Carvalho et al. 2015 for evidence that French 3-year-olds have acquired a certain number of homophone pairs). An important open issue then is to equip the learning system with the right built-in constraints. At this point it seems that word learning is guided by a) learners’ expectations that concepts are convex, always; b) learners’ expectation that word forms are linked to one meaning, in general; and c) the possibility that a word form maps onto several distinct meanings if a) is challenged. While a) helps to constrain the possible concepts one can entertain, b) helps to constrain the number of hypotheses children need to consider when learning words (especially while used in combination with a)). Note that this would constitute very specific priors for a general learning system: it presupposes that children already know that an essential feature of their to-be-learned lexicon is to be composed of form-meaning mappings of very specific kinds. The present study contributes to point c) above: children can entertain the possibility that a word maps onto several distinct meanings to accommodate apparent violations of a) concept convexity at the detriment of b) a one-to-one mapping between forms and meanings. This suggests that children are able to selectively trigger or silence b) as a function of the learning situation (and we documented that this could be made possible by observing the sampling distribution of the learning exemplars). One possibility for current word learning accounts would be thus to implement these three built-in constraints directly in the learning system and tweak the learning inference component to accommodate the sampling effect we document. Yet this would be a rather strong assumption that amounts to saying that the learning system expects the existence of a phenomenon as specific as homophony, from the start.

Because current word learning accounts specialized into fairly simple word learning phenomena (i.e., one form associated with a single meaning), they equipped the learner with specialized built-in constraints (e.g., a one-to-one form-meaning mapping constraint) that cannot explain learning of other more complex phenomena, such as homophony. Incorporating homophony, and potentially other less trivial word learning situations, will allow these accounts to delineate the more general priors that children bring into the word learning task. Yet, although adding different built-in constraints to these accounts may be technically easy, this may be theoretically challenging as we described above, as one would need to explain how the learning system is capable of juggling between different priors to appropriately learn different types of words.

## **Conclusion**

Homophony presents a challenge to word learners: it requires children to discover that word forms and concepts are not always in a one-to-one relation, an otherwise important assumption to restrict the search space for word meanings. The present study showed that children use information about the sampling distribution of learning exemplars (and in particular the fact that they form two distinct convex clusters in conceptual space) to infer homophony for a novel label. We argue that this unexplored sensitivity and the very possibility of homophony should be incorporated into future accounts of word learning.

## **References**

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using eigen and s4 [Computer software manual]. (R package version 1.1-2)



- Beveridge, M., & Marsh, L. (1991). The influence of linguistic context on young children's understanding of homophonic words. *Journal of Child Language*, 18(02), 459–467.
- Bloom, P. (2001). Précis of How children learn the meanings of words. *Behavioral and Brain Sciences*, 24(06), 1095–1103.
- Campbell, R. N., & Bowe, T. (1977). Functional asymmetry in early language understanding. *Salzberger Beiträge für Linguistik*, 3.
- Casenhiser, D. M. (2005). Children's resistance to homonymy: an experimental study of pseudohomonyms. *Journal of Child Language*, 32(2), 319–343.
- Choi, Y., & Trueswell, J. C. (2010). Children's (in) ability to recover from garden paths in a verb-final language: Evidence for developing control in sentence processing. *Journal of Experimental Child Psychology*, 106(1), 41–61.
- Dautriche, I., & Chemla, E. (submitted). What homophones say about words.
- de Carvalho, A., Dautriche, I., & Christophe, A. (2015). Preschoolers use phrasal prosody online to constrain syntactic analysis. *Developmental science*.
- Doherty, M. J. (2004). Children's difficulty in learning homonyms. *Journal of Child Language*, 31(1), 203–214.
- Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science*, 20(5), 578–585.
- Goodman, N. (1955). *Fact, fiction, and forecast*. Harvard University Press.
- Gärdenfors, P. (2004). *Conceptual spaces: The geometry of thought*. MIT press.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434–446.
- Khanna, M. M., & Boland, J. E. (2010). Children's use of language context in lexical ambiguity resolution. *The Quarterly Journal of Experimental Psychology*, 63(1), 160–193.
- Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive development*, 3(3), 299–321.
- Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. MIT Press.
- Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive psychology*, 16(1), 1–27.
- 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.
- Mazzocco, M. M. (1997). Children's interpretations of homonyms: a developmental study. *Journal of Child Language*, 24(02), 441–467.
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological review*, 92(3), 289.

- Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2010). Broca's area and language processing: Evidence for the cognitive control connection. *Language and Linguistics Compass*, 4(10), 906–924.
- Piantadosi, S. T., Tenenbaum, J. B., & Goodman, N. D. (2012). Bootstrapping in a language of thought: A formal model of numerical concept learning. *Cognition*, 123(2), 199–217.
- Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. The MIT Press.
- Plunkett, K., Hu, J.-F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition*, 106(2), 665–681.
- Quine, W. V. O. (1960). *Word and Object*. MIT Press.
- R Core Team. (2013). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria.
- Rabagliati, H., Pyllkkänen, L., & Marcus, G. F. (2013). Top-down influence in young children's linguistic ambiguity resolution. *Developmental Psychology*, 49(6), 1076–1089.
- Regier, T. (2005). The emergence of words: Attentional learning in form and meaning. *Cognitive science*, 29(6), 819–865.
- Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, 61(1), 39–91.
- Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. *Studies of child language development*, 1, 75–208.
- Slobin, D. I. (1975). *Language change in childhood and in history*. Language Behavior Research Laboratory, University of California.
- Srinivasan, M., & Snedeker, J. (2011). Judging a book by its cover and its contents: The representation of polysemous and homophonous meanings in four-year-old children. *Cognitive Psychology*, 62(4), 245–272.
- Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, 114(2), 245–272.
- Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, 18(5), 414–420.
- Zwicky, A., & Sadock, J. (1975). Ambiguity tests and how to fail them. *Syntax and semantics*, 4(1), 1–36.