Cognitive Psychology 104 (2018) 83–105

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

Learning homophones in context: Easy cases are favored in the lexicon of natural languages

Isabelle Dautriche[a,*], Laia Fibla[d], Anne-Caroline Fievet[b,c], Anne Christophe[b,c]

[a] School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh, United Kingdom
[b] Laboratoire de Sciences Cognitives et Psycholinguistique, DEC-ENS/EHESS/CNRS, Ecole normale supérieure – PSL Research University, Paris, France
[c] Maternité Port-Royal, AP-HP, Faculté de Médecine Paris Descartes, France
[d] University of East Anglia, School of Psychology, UK

ARTICLE INFO

Keywords:
Word learning
Homophones
Lexicon
Language acquisition
Lexical design

ABSTRACT

Even though ambiguous words are common in languages, children find it hard to learn homophones, where a single label applies to several distinct meanings (e.g., Mazzocco, 1997). The present work addresses this apparent discrepancy between learning abilities and typological pattern, with respect to homophony in the lexicon. In a series of five experiments, 20-month-old French children easily learnt a pair of homophones if the two meanings associated with the phonological form belonged to different syntactic categories, or to different semantic categories. However, toddlers failed to learn homophones when the two meanings were distinguished only by different grammatical genders. In parallel, we analyzed the lexicon of four languages, Dutch, English, French and German, and observed that homophones are distributed non-arbitrarily in the lexicon, such that easily learnable homophones are more frequent than hard-to-learn ones: pairs of homophones are preferentially distributed across syntactic and semantic categories, but not across grammatical gender. We show that learning homophones is easier than previously thought, at least when the meanings of the same phonological form are made sufficiently distinct by their syntactic or semantic context. Following this, we propose that this learnability advantage translates into the overall structure of the lexicon, i.e., the kinds of homophones present in languages exhibit the properties that make them learnable by toddlers, thus allowing them to remain in languages.

1. Introduction

Carey (1978) has described word learning as starting with a process where children *“flag “new word!” upon hearing a phonological sequence with no current lexical entry“* (p. 272). Indeed, one feature of novel words is that they are often composed of unfamiliar sequences of sounds. Yet that does not have to be always the case: For instance, a child may already know that ”bat“ means bat-animals and be confronted with a sentence such as ”aluminum bats are much easier to swing when compared to wooden bats“. How does the child determine that ”bat“ is used here to refer to a baseball-bat and not an animal-bat? Homophones, whereby a single phonological form is associated to several meanings, thus present learners with a unique word learning situation where they cannot rely on the signal alone to determine whether a phonological form is a candidate for entering the lexicon as a novel word.

Children have well-documented difficulties in learning homophones. Previous research showed that preschoolers perform poorly on tasks requiring them to assign a different, unrelated meaning to a known word form (e.g., learning that “snake” could also refer to a novel object that is not a snake) compared to learning a novel meaning for a novel word form (e.g., learning that “blicket” refers to a

* Corresponding author at: Center for Language Evolution, University of Edinburgh, UK.
E-mail address: isabelle.dautriche@gmail.com (I. Dautriche).

https://doi.org/10.1016/j.cogpsych.2018.04.001
Accepted 26 April 2018
0010-0285/

novel object) (Casenhiser, 2005; Doherty, 2004; Mazzocco, 1997). This suggests that the existing word interferes with the learning of a novel meaning for the word form. This could happen for two reasons: either children fail to realize that a novel word has in fact been presented, and therefore do not even attempt to attribute a second meaning to the well-known word form; or, children do attempt the construction of a novel lexical entry, but they do not succeed in attributing it a meaning, because of interference from the existing meaning already attached to that word form. In relation to children's difficulty with homophones, it has been proposed that children start with the assumption that a word form maps onto exactly one meaning (Slobin, 1985). Such a simplistic assumption on the lexicon predicts learning success in most learning circumstances where word forms and meanings are aligned, and failures in those cases that depart from this simplistic scenario, such as homophones.

Because homophones are difficult to learn, we expect that they should be absent or dispreferred in languages. In line with this idea, several studies revealed some diachronic pressures *against* homophony (André, 1955; Barkal, 1978; Li & Thompson, 1989; Wedel, Kaplan, & Jackson, 2013): For example, Wedel et al. (2013) showed that two sounds are less likely to merge if they result in a larger amount of homophony in the language. Yet, despite the presence of such pressures, and of children's learning difficulties, homophony is a common occurrence across languages (about 4% of word forms are homophones, see Dautriche, 2015), especially among shorter words which are the most frequently used part of the lexicon (Piantadosi, Tily, & Gibson, 2012). This fact seems to challenge theories arguing that the properties of language are shaped by biases and limitations on human cognitive systems (e.g., Bates & MacWhinney, 1982; Chomsky, 1965; Christiansen & Chater, 2008; Slobin, 1978).

The present work addresses this apparent discrepancy between learning abilities and typological pattern, with respect to homophony in the lexicon. We propose that (a) learning homophones is easier than previously thought, at least when the meanings of the same phonological form are sufficiently distinct; (b) this learnability advantage translates into the overall structure of the lexicon, i.e., the kinds of homophones present in languages exhibit the properties that make them learnable by toddlers, thus allowing them to remain in languages.

In particular, we propose that children's ability to learn homophones may crucially depend on the context they are presented in. Indeed, words are rarely uttered in isolation but are part of the broader context in which they are used: the sentence in which they are pronounced, the discourse, the speakers involved, the register of language, the surroundings, etc. Adults constantly form linguistic expectations based on the linguistic and extra-linguistic contexts (Altmann & Kamide, 1999; Creel, Aslin, & Tanenhaus, 2008; Federmeier & Kutas, 1999; Millotte, René, Wales, & Christophe, 2008; Nieuwland & Van Berkum, 2006; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). As a result, lexical access may be sufficiently constrained, with the consequence that homophones may not be detrimental for everyday speech comprehension, so long as only one member of a homophone pair is expected in a given context. The fact that adults hardly notice homophones proves the efficiency of such a context-dependent language processing system. Similarly, learning a novel meaning for a known word form may be easier if the novel and the known meanings appear in distinct contexts, and if children can capitalize on these different contexts. For instance, homophones may be easier to learn when their meanings are sufficiently distant syntactically (e.g. "an eat" may be a good label for a novel animal), or semantically (e.g. "a glass" as a new label for a novel animal), but not when they are close (e.g. "a cat" for a novel animal). In other words, presenting a known word form in a context that is distinct from its original use may eliminate the possibility that the original meaning was intended, and thereby boost the likelihood that a novel meaning is introduced.

In a series of experiments, we manipulated several sources of information that may help children to identify when they should postulate a novel meaning for a known word form. We manipulated the syntactic and/or the semantic distance between the known and the novel meaning, as well as the neighborhood density of the phonological form of the word. In these experiments, French 20-month-old toddlers were taught novel words that are homophonous to familiar words. We tested 20-month-olds, thus younger children than the population tested in previous studies testing homophone learning (Casenhiser, 2005; Doherty, 2004; Mazzocco, 1997), for several reasons: (1) Toddlers of this age already use the different sources of information that we propose to investigate here (noun vs. verb, e.g., Cauvet et al., 2014; semantic relations, e.g., Arias-Trejo & Plunkett, 2013; gender cues, e.g., Van Heugten & Christophe, 2015; neighborhood density, e.g., Newman, Samuelson, & Gupta, 2008) and (2) 20-month-olds have already acquired a certain number of homophone pairs (see de Carvalho, Dautriche, & Christophe, 2017) suggesting that learning lexical ambiguities should be possible at this young age.

We then used the results of these experiments to evaluate whether the dimensions that make homophones easier to learn have a visible impact on the distribution of these words in the lexicon: Specifically, we tested whether homophone pairs that are learnable by toddlers are over-represented in the lexicon of natural languages, relative to homophone pairs that are harder to learn (using 4 languages as a test case).

2. Experimental studies

2.1. Experiment 1 – syntactic and semantic distance

Words from different syntactic categories typically occur in different syntactic contexts (e.g., "You X the" is a felicitous context for a verb but not for a noun) and children have been shown to use such contextual information to recognize known words by the age of 18 months (e.g., Brusini et al., 2017; Cauvet et al., 2014; Kedar, Casasola & Lust, 2006; Zangl & Fernald, 2007). Learning a pair of homophones may thus be easier if the two members of the pair can be distinguished syntactically. One way to achieve this is to distribute members of a pair of homophones in different syntactic categories, such that the homophones will occur in different syntactic contexts. For instance, in English, "to train", a verb, which designates an action, is homophonous with "train", a noun, which designates an object. These two words, although they have the same phonological form, are thus distinct on the syntactic dimension (noun/verb); in addition, they also differ in the semantic dimension (action/object) – we will come back to this point in the Discussion.

Previous studies found that 18-month-old toddlers find it difficult to learn a novel word, when that word differs from a known word by a single phonological distinction: They do not manage to learn that a novel animal is called a "tog", a phonological neighbor of the known noun "dog" (Dautriche, Swingley, & Christophe, 2015; Swingley & Aslin, 2007). However, this difficulty disappears when the novel word appears in a syntactic context that is distinct from the one of its familiar competitor (e.g. learning that a novel animal is called a "kiv", a phonological neighbor of the known verb "give"; Dautriche et al., 2015). In addition, Casenhiser (2005) found that 4-year-olds find it easier to learn an additional meaning for a known word, thus a homophone, when it is used in a different syntactic context (learning that "a give" could label a novel object). Experiment 1 seeks to extend these results to 20-month-old French toddlers learning novel homophones.

Toddlers were taught an animal label that is homophonous with a known verb (a **verb-homophone**, e.g., "un manger" *an eat*) and another animal label that has no homophone in their lexicon (a **non-homophone**, e.g., "un torba"). If toddlers take into account syntactic or semantic likelihoods when identifying whether a given word form instantiates a novel meaning rather than a word they know, they should be able to learn both verb-homophones and non-homophones (taught as animal labels).

2.1.1. Method

Word learning was evaluated using a preferential looking method (Fernald, Zangl, Portillo, & Marchman, 2008). An above-chance proportion of looks towards the target picture after word onset was taken as evidence that the word had been learned.

All the procedure, design, criterion for rejection and analysis plan were decided prior to data collection and pre-registered here https://osf.io/ab63r/?view_only=8ec89a0430d5453fa28ec2d80c6bc888 for Experiment 1 and here for Experiment 2 and 5: https://osf.io/pyz8x/?view_only=721d951af31d4b40be068a0b6fa746e9. Experiments 3 and 4 used the same paradigm and criteria but were not pre-registered. Note that for Experiments 2–5, we changed (minimally) the design to rule out an alternative interpretation of our results (see supplemental material). Nevertheless, the original experiments are reported in the supplemental material. Stimuli, script for analyses and participants' results are accessible through https://osf.io/pyz8x/?view_only=721d951af31d4b40be068a0b6fa746e9.

Participants. Thirty-two monolingual French 20-month-olds, ranging from 19;1 to 20;9 with a mean age of 20;3, (SD = 0;5; 13 boys) took part in this experiment. Five additional children were replaced because of fussiness during the experiment resulting in more than 50% of trials with missing eye tracking data (n = 3), experimenter error (n = 1), born at less than 37 weeks' gestation (n = 1) (see below criteria for trials and participants exclusion).

Criteria for trial and participant exclusion. Trials for which we have less than 75% of eye-data were rejected. Children having less than 50% of valid test trials in each condition were removed from the analysis.

In addition, we discarded data for children who did not show any preference (more than 55% of looks) for the target picture for familiar words (based on the average proportion of looks for all valid familiar trials, from the onset of the target word until the end of the trial). This was to ensure that we keep only children who are on task (only one participant was excluded by this criterion in the set of experiments we present in this paper – the word-recognition task is generally very easy for toddlers of this age).

Apparatus, procedure and design. In Experiment 1, each child was taught two novel words, a verb-homophone (a known verb in a noun frame to label a novel object, e.g. "an eat") and a non-homophone (both described in detail in the section below).

Before coming to the lab, parents filled out a vocabulary questionnaire that included all the tested verb-homophones and all their phonological neighbors (as we controlled for the phonological neighborhood density of these words in children's lexicon). Each toddler was taught a novel meaning for a label (s)he already knew (according to her/his parents). Toddlers who didn't know any of the 4 verbs were rejected from the experiment. Toddlers sat on their parent's lap about 70 cm away from the screen (27-inch monitor) in a soundproof test cabin. Their eye movements were recorded by an Eyelink 1000 eye-tracker. We used a 5-point calibration procedure. Once the calibration was judged acceptable by the experimenter, the experiment began.

The experiment included two phases: a teaching phase and a testing phase.

Teaching phase. During the teaching phase, children first saw an introductory video that showed a woman playing with a car ("une voiture") and labelling it several times in a short story. This was to ensure that children got familiar with the procedure, and understood that the speaker would talk about the objects she was manipulating. Then they saw 4 teaching videos (2 for each novel word). The objects that were used as the referents of the verb-homophone and the non-homophone were counterbalanced across children. The order of presentation of the teaching videos was interleaved between the two words and counterbalanced across toddlers. After the teaching videos were presented, the test phase began as soon as children looked at a fixation point displayed on the screen.

Test phase. The test phase consisted of 16 trials: 8 filler trials with familiar words and 8 test trials with novel words (4 per novel word). Each trial started with the simultaneous presentation of two pictures on the right and left sides of the screen. Two seconds later, the audio stimuli started: ("*Regarde le [target], tu le vois le [target]?*" Look at the [target], do you see the [target]?). The trial ended 3.5 s after the first target word onset. Trials were separated by a 1 s pause. Immediately consecutive trials did not contain the same target word; there were no more than 2 test trials in a row. Target and distractor pictures appeared equally often on the right and on the left side of the screen across the test phase. Target side did not repeat in more than two consecutive trials. The whole experiment lasted about 5 min.

Materials. *Target words.* To ensure that toddlers would be tested on a verb label they already have a meaning for, we selected 4 phonological word forms of verbs that toddlers of that age are likely to know (according to the French adaptation of the MacArthur-Bates CDI, Kern, 2007, collected in previous studies), as well as 4 novel word forms.

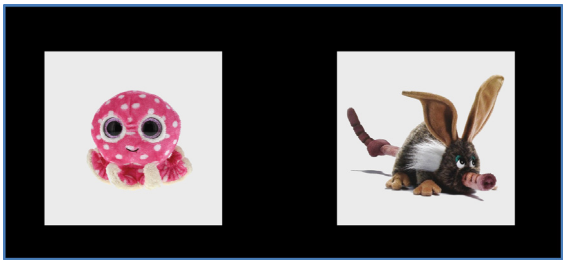

Fig. 1. Target objects used in Experiment 1, 2, 4 and 5.

The 4 non-homophones were all bisyllabic: “prolin”, “barlier”, “torba”, “lagui” (/pʁolɛ̃/, /barljé/, /tɔʁba/, /lagi/) and they had no phonological neighbors within the toddlers’ vocabulary. All 4 words were used with a masculine gender as their endings were predominantly masculine: As calculated on a French lexical database (New et al., 2004), the percentage of masculine bisyllabic words ending in /ɛ̃/ is 98%, ending in /jé/, 95%, ending in /a/, 76% and ending in /i/, 73%. The 4 verb-homophones were also all bisyllabic words: “manger”, “tomber”, “casser”, “cacher” (/mɑ̃ʒe/, /tõbe/, /kase/, /kaʃe/) meaning: *eat*, *fall*, *break* and *hide*. The verb-homophones were the infinitive and the past participle forms of the known verbs (which are identical, in French, for this set of verbs). These forms were used because they were bisyllabic and they are very common (the most frequent morphological form for all 4 verbs; Demuth & Tremblay, 2008). The average duration of the novel words in the test sentences was 640 ms for the verb-homophones and 622 ms for the non-homophones. The difference between target words durations was not significant.

Target objects. The target objects were two unfamiliar animals. One resembled a pink white-spotted octopus with an oversized head. The other looked like a rat with bunny ears and a trunk (see Fig. 1).

Teaching videos. Word teaching was done though several video recordings presented on a television screen. The teaching phase included four short videos of about 30 s each. In each video, the speaker talked about the target object she was playing with and labelled it 5 times using one of the target words. The verb-homophone was used in two videos and the non-homophone word in the other two. In total, toddlers heard each novel word 10 times.

Testing stimuli. Our visual stimuli were photographs of objects on a light gray background. For the familiar trials, we chose 8 objects that children of that age are likely to know: *voiture, banane, poussette, chaussure, chien, poisson, cuillère, maison* (car, banana, baby-stroller, shoe, dog, fish, spoon, house). Pictures were yoked in gender-matched pairs (e.g. the banana, une$_{fem}$ banane$_{fem}$, always appeared with the car, une$_{fem}$ voiture$_{fem}$). For test trials, the pictures of the two target animals were always presented together.

The audio stimuli consisted of the sentences “*Regarde le [target], tu vois le [target]?”* Look at the [target]. Do you see the [target]?) or “*il est où, le [target]? Regarde le [target].”* (Where is the [target]? Look at the [target]!) were [target] was the tested word; the target word was thus pronounced twice in each test trial. All sentences were recorded by a native French speaker (the last author, and the same speaker as in the videos).

Measurement and analysis. Gaze position on each trial was recorded via an eye-tracker with a 2 ms sampling rate. For a few children, the sampling rate was 4 ms due to some changes in the eye-tracker settings. For all analyses and plots, we coded the proportion of target looks relative to total looks to target and distractor (excluding looks away) for each 50 ms time bin. The time course of eye movements was inspected from the beginning of the first target word (“Look at the [target]”) until the end of the trial. Since no previous work looked at homophone learning in a preferential looking task, we were unsure as to whether increased looks towards the target should start at about the same time as for familiar words (about 500 ms after the beginning of the target word according to previous studies with children of about the same age; see Dautriche et al., 2015), so we considered the whole trial duration in our analysis.

We conducted three cluster-based permutation analyses (as in Dautriche et al., 2015; de Carvalho, Dautriche, & Christophe, 2016; de Carvalho et al., 2017; Hahn, Snedeker, & Rabagliati, 2015; Von Holzen & Mani, 2012; see Maris & Oostenveld, 2007 for a formal presentation of the analysis), a non-parametric statistical test: each condition (verb-homophone; non-homophone) was compared to chance by comparing the average proportion of looks towards the target picture to 50% (the chance level); and one analysis compared the looking proportions between conditions.

The cluster-based permutation analysis works in the following way: at each time point we computed a *t*-test to compare fixations to the target compared to chance (0.5) or to compare fixations to the target between the two experimental conditions (depending on the analysis). All fixation proportions were transformed via the arcsin square function to fit better the assumptions of the *t*-test. Adjacent time points with a significant effect ($t > 2$; $p < .05$) were grouped together into a cluster. The size of each cluster was computed as the sum of all t-values within this cluster.

To obtain the probability of observing a cluster of the same size by chance, we conducted 1000 simulations where conditions (verb-homophone vs. chance, non-homophone vs chance, or verb-homophone vs. non-homophone) were randomly assigned for each trial. For each simulation, we computed the size of the biggest cluster identified with the same procedure that was used for the real data. A cluster from the original data was considered significant if the probability of observing a cluster of the same size or bigger in the randomized data was smaller than 5%, corresponding to a p-value of 0.05.

It should be mentioned that the criterion for including a time bin in a cluster ($t > 2$ in the present study) is independent of the process which assesses cluster significance, and thus does not affect the likelihood of a false positive. However, it does influence the size of the time window that one can find; if the threshold is lower then the time-window will be wider. Yet, the same threshold is applied to the randomized data as well, such that the chance of getting a bigger cluster will also increase under the null hypothesis, maintaining the rate of false positives under 0.05.

2.1.2. Results

Fig. 2 shows the proportion of looks towards the target picture, time-locked to the beginning of the first target word ("Regarde le **[target]** ! tu le vois le [target]" *Look at the **[target]**! Do you see the [target]*) for the whole duration of the trial.

Toddlers learnt both the verb-homophone (green curve) and the non-homophone (black curve). The cluster-based permutation analysis (Maris & Oostenveld, 2007) identified a significant time-window where toddlers fixated the target picture significantly above chance (0.5) when asked to look at the verb-homophone (from 1500 ms to 2600 ms after word onset, $p < .01$; green-shaded time-window in Fig. 2) and when asked to at the non-homophone (from 1100 ms to 1850 ms and from 2800 ms to 3500 ms after word onset, $p_s < .05$; gray-shaded time-windows in Fig. 2).[1] There was no significant difference between conditions.

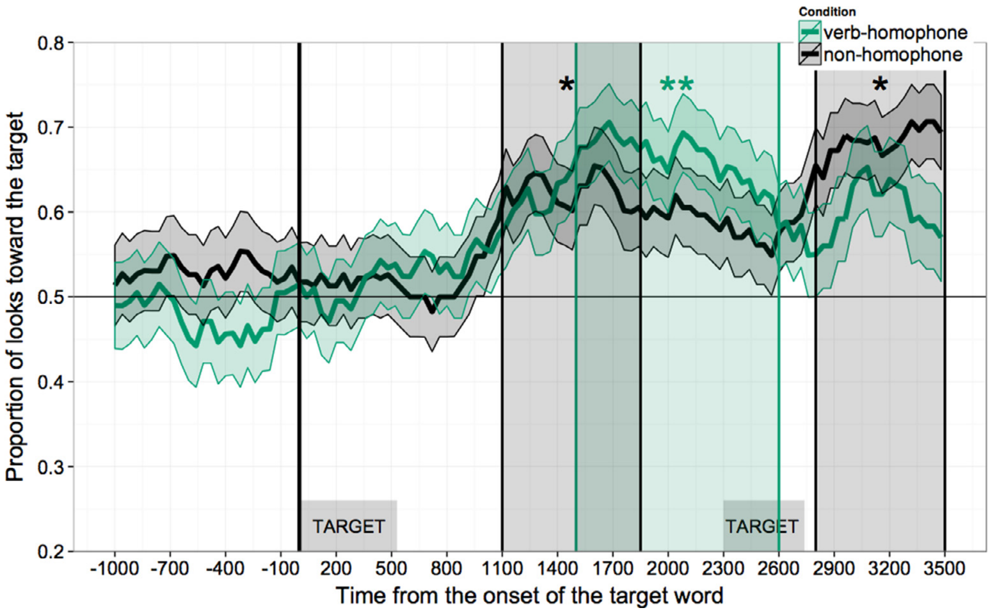

Fig. 2. Proportion of looks towards the target picture, time-locked to the beginning of the first target word (do you see the [target]? where is the [target]? (the gray boxes at the bottom of the plot represent the position, in time, of the two target words within the test sentences during the course of a trial) for the verb-homophone (in green) and for the non-homophone (in black). The ribbon surrounding each curve represents the standard error of the mean obtained at each time bin for each word condition. Toddlers successfully learnt both the verb-homophone and the non-homophone, as evidenced by the fact that they significantly increased their looks towards the correct picture after the target word was pronounced, in both conditions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.1.3. Discussion

Toddlers successfully learnt the verb-homophone and the non-homophone. Alternatively, it may be the case that toddlers only learnt the non-homophone and simply looked at the other object when hearing the verb-homophone (a process known as mutual exclusivity; Markman & Wachtel, 1988). Yet, we find this possibility unlikely. In another study (Dautriche et al., 2015), where we similarly taught toddlers two novel words, toddlers failed to apply mutual exclusivity: They successfully learnt one of the novel words but failed to recognize the other. This suggests that in the present case, toddlers had no problem learning a verb-homophone. Our subsequent experiments will address that potential concern more directly as only the homophone label will be taught, ruling out the use of mutual exclusivity from the start.

Success in learning a verb-homophone may be surprising given that several studies have shown that even preschoolers find it difficult to associate a novel meaning to a known word form (Casenhiser, 2005; Doherty, 2004; Mazzocco, 1997). Yet contrary to previous studies, toddlers may not have been disturbed by the re-use of a known verb, as the syntactic context, a noun frame, decreased the likelihood that the speaker could possibly be using the known meaning of the verb (see also Casenhiser, 2005 for a similar result in 4-year-olds). This suggests that children can use the syntactic context to identify whether a novel meaning is appropriate for a word, even when the word is already associated to a meaning in the child's lexicon.

[1] The fact that we are observing two significant time-windows for the non-homophone condition, rather than a single long time-window, depends on the parameters of the cluster-based analysis: Because the proportion of looks towards the non-homophone target decreases around 2000 ms, in-between the two sentences asking the child to look at the target, the t-value of the comparison between proportion of looks at target and the chance level (0.5) falls below the pre-defined threshold (t = 2), which yields two separate clusters. We could obtain a single larger cluster collapsing these two by decreasing the t threshold, yet to be consistent with the other analyses, we left the threshold as we defined it in the method section (see also Footnote 3).

One possibility is that toddlers did not even notice that a familiar verb was used in this experiment. Indeed, toddlers use the syntactic context to constrain lexical access: For instance, they expect a noun after a determiner and a verb after a pronoun (e.g., Cauvet et al., 2014; Shi & Melancon, 2010; Zangl & Fernald, 2007). Thus, when presented with a noun phrase such as "C'est un manger!" *this is an eat!*, toddlers may simply have failed to access the familiar verb, which would make it very easy for them to link an additional meaning to the word, just as easy as for the non-homophone.

Another possibility is that toddlers initially noticed that the verb they knew was used in an incorrect frame. Indeed, two-year-old children display an early left-lateralized brain response when an expected noun is incorrectly replaced by a verb (e.g., "je prends la mange" *I take the eat*, (Bernal, Dehaene-Lambertz, Millotte, & Christophe, 2010, Brusini, Dehaene-Lambertz, Dutat, Goffinet, & Christophe, 2016; Brusini et al., 2017). Relevantly, ERP studies looking at adults' processing of lexical ambiguities in reading tasks found that the syntactic context alone was insufficient to constrain lexical access in the case of noun/verb homographs (e.g., the park/ to park) as evidenced by a frontal negativity compared to unambiguous words (e.g., Lee & Federmeier, 2009). Yet additional semantic constraints on the meaning of the words eliminated the frontal negativity (e.g., Lee & Federmeier, 2009). Such a frontal component has been suggested to reflect the recruitment of frontally mediated meaning selection mechanisms needed to disambiguate noun-verb homographs in the absence of constraining semantics (e.g., Novick, Trueswell, & Thompson-Schill, 2010) suggesting that the syntactic context alone may be insufficient to suppress totally the inappropriate meaning of the word. Yet, in our experiment, because the verb is never used to convey its initial meaning (thus the verb meaning is never pre-activated) but repetitively used in a noun frame with a supporting visual context (i.e., the novel referent), social support (i.e., the speaker looking contingently to the referent each time she uses the verb-homophone) and supplemented with information about its novel meaning (e.g., "Un manger, ça a des grandes oreilles" *Eats have big ears*), this may have increased the likelihood that the known meaning of the verb was inappropriate in that context, and supported the identification of an additional meaning for the known verb form.

At any rate, we showed that learning homophones may be easier when their meanings are made sufficiently distinct by the context in which they are used. When a *verb* form referring to an *action* was used as a *noun* to label a novel *animal* (or an object more generally), children had no problem to learn the link between the known phonological form and the novel meaning. This result fits well with the observation that noun/verb homophones are already known by very young children (de Carvalho et al., 2016, 2017; Veneziano & Parisse, 2010) and thus do not present major acquisition difficulties for children. Yet our results cannot tell whether this effect is due to the syntactic distance (verb/noun), or to the semantic distance (action/object), or both. The following set of experiments investigates the effect of semantic distance alone, when teaching toddlers noun-noun homophones.

2.2. Experiment 2 – semantic distance

Learning homophones may be easier when the semantic distance between their meanings is large. For example, "bat" is likely to be unambiguous in a context where one speaks about sport, as we do not expect the bat-animal meaning in this context. Intuitively, homophones seem to map onto clearly distinct meanings (e.g., animal-bat/baseball-bat, flour/flower, mussel/muscle, etc.) suggesting an advantage for homophones that are semantically distinct over semantically close.

In Experiment 2, toddlers were taught either a noun for a novel animal that is homophonous with a noun referring to an artifact (**an artifact-homophone**, syntactically identical, but semantically distant from the to-be-learnt meaning; e.g., "un pot", *a potty*) or a noun for a novel animal that is homophonous with a noun referring to an animal (**an animal-homophone**, both syntactically and semantically close; e.g., "un chat", *a cat*). Contrary to Experiment 1, children cannot use the syntactic context to distinguish between the two meanings of the homophone (as they both appear in a noun syntactic context). Yet, if toddlers evaluate the semantic features of words when identifying novel words, artifact-homophones, but not animal-homophones, should be perceived as sufficiently distinct from their familiar competitor to be assigned a novel meaning.

Contrary to Experiment 1, only one target word was taught to children in this experiment, so as to discard the possibility that children could use mutual exclusivity to succeed in this task (see Discussion of Experiment 1): one group of children was taught an artifact-homophone; another group was taught an animal-homophone. All the children also saw videos of an unnamed animal. Because children were taught only one word, we also controlled that a potential learning effect would not be due to a preference for looking at the only labelled animal: in other words, at test, toddlers might look more towards the labelled animal not because they associated the novel label to this animal, but because this animal is the only one that has been named during the teaching phase. Therefore, during the test phase, toddlers were tested on the homophone word (the artifact-homophone or the animal-homophone, depending on their group), and on an untaught non-homophone. When tested on the non-homophone, if toddlers look more towards the unlabeled animal (or at least behave differently than when tested on the homophone word, as expected if they follow mutual exclusivity, Markman & Wachtel, 1988), this would suggest that toddlers learnt the form-meaning association between the artifact-homophone (or the animal-homophone) and its animal referent.

2.2.1. Method

Participants. Thirty-two French 20-month-olds took part in this experiment, sixteen learnt an artifact-homophone (range = 19;2 months to 20;8 months, mean = 19;8, SD = 0;6, 8 boys) and sixteen learnt an animal-homophone (range = 19;4 months to 20;9 months, mean = 20;2, SD = 0;4, 6 boys). Sixteen additional children were replaced because of technical problem (n = 7),[2] fussiness

[2] For this Experiment, the position of the eye-tracker relative to the screen was improperly centered, resulting in an increased likelihood to lose data when children were looking towards one side of the screen.

during the experiment resulting in more than 50% of trials with missing eye tracking data (n = 6), refusal to wear the sticker necessary for eye-tracking (n = 2) and not knowing any of the test words according to their parents' report (n = 1).

Apparatus, procedure and design. Similar to Experiment 1, except that we taught only one novel word to the participants (either an artifact- or an animal-homophone). Participants still saw 5 videos: one training video, 2 teaching videos with one of the novel animal named with either the artifact-homophone or the animal-homophone label, and 2 videos with the other novel animal, that respected the same movements and story line of the 2 other videos, but without any label for the novel animals (only pronouns were used – e.g. "do you see this one? It has big ears…"). During the test phase, toddlers were presented with 4 test trials on the homophone label (the artifact- or the animal-homophone) and 4 test trials on a non-homophone label (the non-homophone condition).

Material. The 4 artifact-homophones were all monosyllabic words: "verre", "pot", "pull", "bain" (/vɛʁ/, /po/, /pyl/, /bɛ̃/) meaning: *glass, potty, sweater* and *bath*. On average, these words had a phonological neighborhood density of 3.8 in children's lexicon (according to the parental report) and an average frequency count[3] of 152 in a corpus of child directed speech (the Lyon corpus, Demuth & Tremblay, 2008).

The 4 animal-homophones were also all monosyllabic words: "chat", "loup", "poule", "mouche" (/ʃa/, /lu/, /pul/, /muʃ/) meaning: *cat, wolf, hen* and *fly*. These words had an average phonological neighborhood density of 3.1 in children's lexicon (according to the parental report) and a frequency count of 157 in a corpus of child directed speech (the Lyon corpus, Demuth & Tremblay, 2008).

There was no statistical differences in frequency nor in phonological neighborhood density between the animal- and the artifact-homophones.

The 4 non-homophones were identical to those used in Experiment 1.

The average duration of the target words in the test sentences was 455 ms for the artifact-homophones, 517 ms for the animal-homophones and 622 ms for the non-homophones.

Measurement and analysis. Similar to Experiment 1, except that this time the dependent variable was the proportion of looks towards the stuffed animal labelled during the learning phase (in order to compare any potential difference in behavior between the trials where the homophone label was tested and the trials where the non-homophone label was tested).

We conducted three cluster-based permutation analyses for each group of children (animal-homophone; artifact homophone): each condition (animal-homophone; non-homophone or artifact-homophone; non-homophone) was compared to chance by comparing the average proportion of looks towards the labelled stuffed animal picture to 50% (the chance level); and one analysis compared the looking proportions between conditions (animal-homophone vs. non-homophone or artifact-homophone vs. non-homophone). In a last analysis we compared the two groups of participants on the conditions of interest (animal-homophone vs. artifact-homophone).

2.2.2. Results

Artifact-homophone group

Fig. 3 shows the proportion of looks towards the referent of the artifact-homophone (the animal that was labelled during the learning phase), time-locked to the beginning of the first target word ("Regarde le **[target]** ! tu le vois le [target]" *Look at the **[target]**! Do you see the [target]*), and plotted from -1500 ms before the beginning of the first target word until the end of the trial.

Toddlers looked towards the target in the artifact-homophone condition, significantly more than chance (blue curve; from 1350 ms to 2550 ms after target word onset; $p < .001$; light blue-shaded time-window in Fig. 3) and they did not show any preference for any of the objects in the non-homophone condition ($p > .3$). Crucially there was a significant difference in performance between the artifact-homophone condition and the non-homophone condition (from 2000 ms to 2500 ms after target word onset; dark blue-shaded area; $p < .05$). This difference in performance ensures that the increase in looking towards the artifact-homophone was not due to a preference for looking at the only labelled animal. See the Supplemental Material for a replication of this effect with another group of 16 toddlers.

Animal-homophone group

Fig. 4 shows the proportion of looks towards the referent of the animal-homophone, time-locked to the beginning of the first target word ("Regarde le **[target]** ! tu le vois le [target]" *Look at the **[target]**! Do you see the [target]*), and plotted from -1500 ms before the first target word until the end of the trial.

Crucially, toddlers behaved differently when they learnt an animal-homophone: In both the animal-homophone condition (gold curve) and the non-homophone condition (black curve) they looked at the animal-homophone referent *before* the onset of the target word, though this came out significant only for the non-homophone condition (from -500 ms to 50 ms around target word; $p < .05$; for the animal-homophone $p = .12$). There was no difference between conditions. Thus, not only did toddlers in the animal-homophone condition fail to show any recognition of the animal-homophone label, they also failed to apply mutual exclusivity when tested on a non-homophone, suggesting that they did not learn to associate the animal-homophone word to the correct stuffed animal.

In a last analysis, we compared the two groups of participants: The difference between the artifact-homophone condition and the animal-homophone condition did not satisfy our criteria for significance ($p = .067$). We speculate that this analysis suffered from a lack of power, since each group behaved clearly differently from one another (see within-group analysis).

[3] These are raw counts in the parents' production.

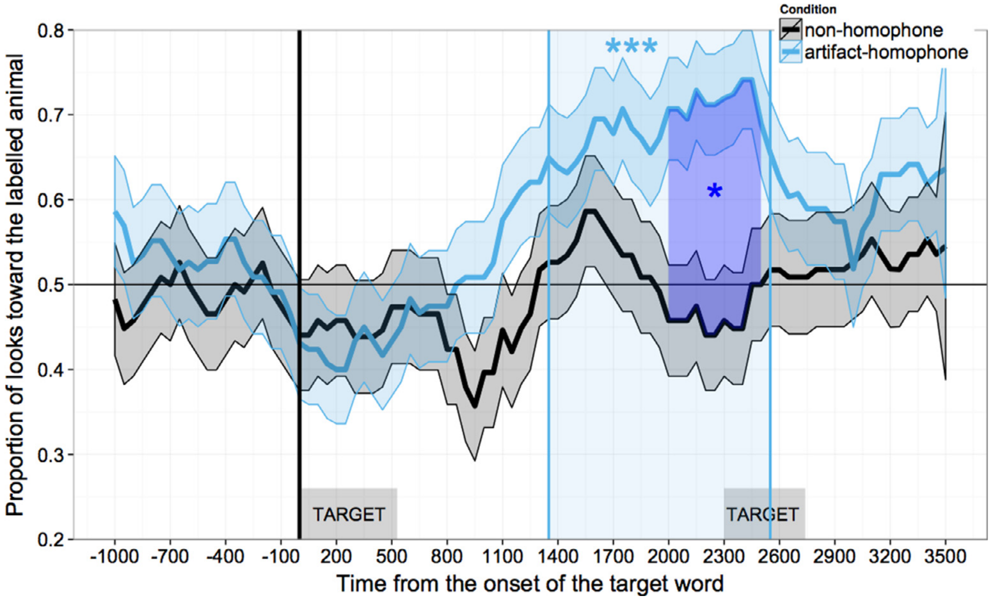

Fig. 3. Result of the artifact-homophone condition. Proportion of looks towards the artifact-homophone referent time-locked to the beginning of the first target word (do you see the [target]? where is the [target]? (the gray boxes at the bottom of the plot represent the position, in time, of the two target words within the test sentences during the course of a trial) for the artifact-homophone (in blue) and for the non-homophone (in black). The ribbon surrounding each curve represents the standard error of the mean obtained at each time bin for each word condition. Toddlers successfully learnt the artifact-homophone, as they significantly increased their looks towards the correct picture after target onset (the light blue-shaded rectangle shows the time-window during which the artifact-homophone condition is significantly different from chance), and this behavior was significantly different from the non-homophone condition, in which children tended to switch back to the unnamed stuffed animal (the dark blue-shaded area shows the time-window in which the two conditions are significantly different from one another). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

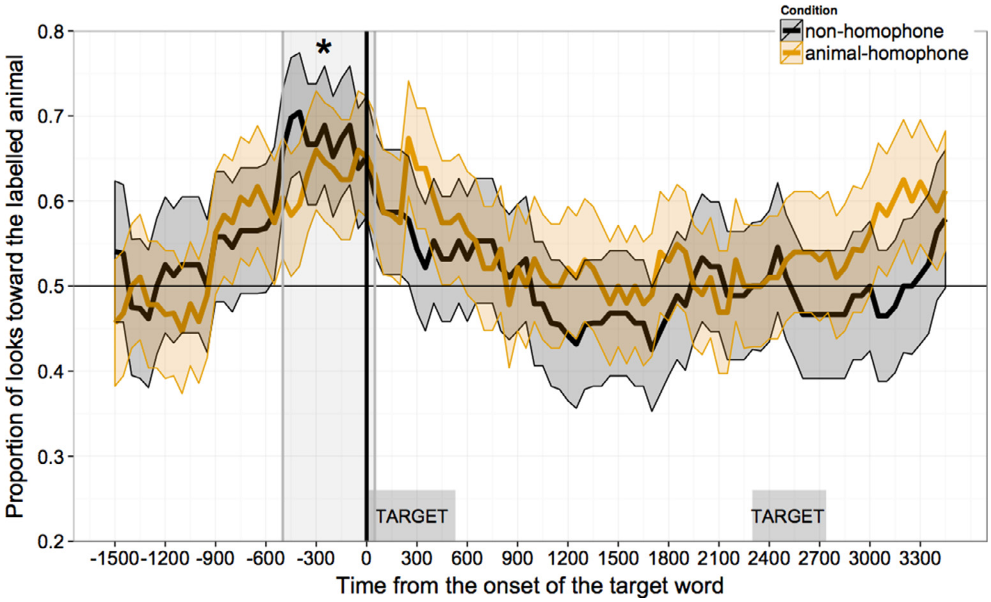

Fig. 4. Results of the animal-homophone condition. Proportion of looks towards the animal-homophone referent, time-locked to the beginning of the first target word (do you see the [target]? where is the [target]? (the gray boxes at the bottom of the plot represent the position, in time, of the two target words within the test sentences during the course of a trial) for the animal-homophone (in gold) and for the non-homophone (in black). The ribbon surrounding each curve represents the standard error of the mean obtained at each time bin for each word condition. Toddlers showed no learning of the animal-homophone: There was no significant increase of looks towards the animal-homophone referent after target word onset in the animal-homophone condition and no difference between the animal-homophone and the non-homophone conditions. Yet toddlers in both conditions tended to look towards the animal-homophone referent *before* target word onset, though this came out significant only for the non-homophone condition (gray shaded area). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Importantly, we replicated a variation of this experiment with another group of 16 toddlers (see the Supplemental Material) and report exactly the same results: Toddlers look at the animal-referent homophone before target word onset, suggesting that this behavior is solid. Yet why would they look at the animal-homophone referent when it appears on the screen? We speculate that they were surprised by such a form-meaning association and tried to accumulate more evidence as soon as they were presented with the animal-homophone referent (e.g., "is this really a cat?"). Note that this assumes that they may actually remember that this toy animal could be labelled "cat". Yet, the fact that they are "surprised" (or behave differently from the artifact-homophone group) is enough for our purpose to say that considerations of semantic distance between the two meanings of a homophone matter during word learning (at least initially).

2.2.3. Discussion

Toddlers had no problem learning an artifact-homophone but failed to display any learning of the animal-homophone. This suggests that toddlers treated these labels differently, and this critically affected their identification of what counts as a novel lexical entry.

One possibility is that they found it easier to learn homophones that are semantically distinct over homophones that are semantically close. When the speaker used an artifact label to name a novel animal, the difference between the normal usage of the word and this novel situation is so great that it looks unlikely that the speaker could use the label to refer to the original meaning. However, when the original meaning (an animal) is close enough to the novel meaning (another animal), as in the case of the animal-homophone condition, it may be more difficult for toddlers to differentiate between a less prototypical member of the original meaning of the label and a novel meaning instance.

Another possibility is that some unmeasured difference between the set of artifact labels and the set of animal labels was responsible for the observed effect. While both sets of words were matched for frequency, neighborhood density in toddlers' lexicon, and phonotactic probability, toddlers may have a better lexical representation for the animal labels than the artifact labels used in this study, leading to greater interference (e.g., Setoh, Wu, Baillargeon, & Gelman, 2013, for some evidence that animals may have a special status). Thus, toddlers may find it more difficult to learn a second meaning for an animal-label than for an artifact-label, not because the semantic distance between the two meanings is greater for the artifact-homophones, but because they have greater difficulty in suppressing the primary meaning of the animal-homophones. The next experiment disentangles between these two possibilities.

2.3. Experiment 3 – Semantic distance

Experiment 3 was similar to Experiment 2, except that this time toddlers were taught the animal-homophones used in Experiment 2 as labels for novel artifacts (e.g., "un chat", a cat, was used to label a novel music instrument). Thus, the set of animal-homophone labels was identical to Experiment 2 but, crucially, the semantic distance between the two meanings of the label increased. If semantic distance between meanings of a pair of homophones is a major reason why learning an artifact-homophone was easier than learning an animal-homophone in Experiment 2, then toddlers should have no problem learning an animal-homophone when the novel meaning is sufficiently distant semantically from the animal category. On the contrary, if better lexical representations for the animal- labels used in Experiment 2 led to the observed effect, then toddlers should still fail to learn a novel meaning for an animal-homophone, even though its second meaning is semantically distinct from its original meaning.

2.3.1. Method

Participants. Sixteen French 20-month-olds took part in this experiment (range = 19;0 months to 20;6 months, mean = 19;6, SD = 0;6, 9 boys). Seven additional children were replaced because of fussiness during the experiment resulting in more than 50% of trials with missing eye-tracking data.

Apparatus, procedure and design. Identical to Experiment 2.

Materials. Similar to Experiment 2 except for the set of target objects used during the teaching phase.

Target objects. The target objects were two unfamiliar artifacts. One was a music instrument composed of 8 spinning coloured bells and the other was a coloured spinning top (see Fig. 5).

Teaching videos. We created 4 new teaching videos of 30 s (2 per object) where the speaker was playing with one of the target objects while labelling it 5 times using the target words. We made sure that the target objects were treated like artifacts: They were the patients of actions but never behaved as agents.

Measure and Analysis. Identical to Experiment 2.

2.3.2. Results

Fig. 6 shows the proportion of looks towards the referent of the animal-homophone time-locked to the beginning of the first target word ("Regarde le **[target]** ! tu le vois le [target]" *Look at the **[target]**! Do you see the [target]*) until the end of the trial.

Importantly, the animal-homophone label did not trigger a "surprisal effect" as in Experiment 2: When the animal-homophone

Fig. 5. Target objects used in Experiment 3.

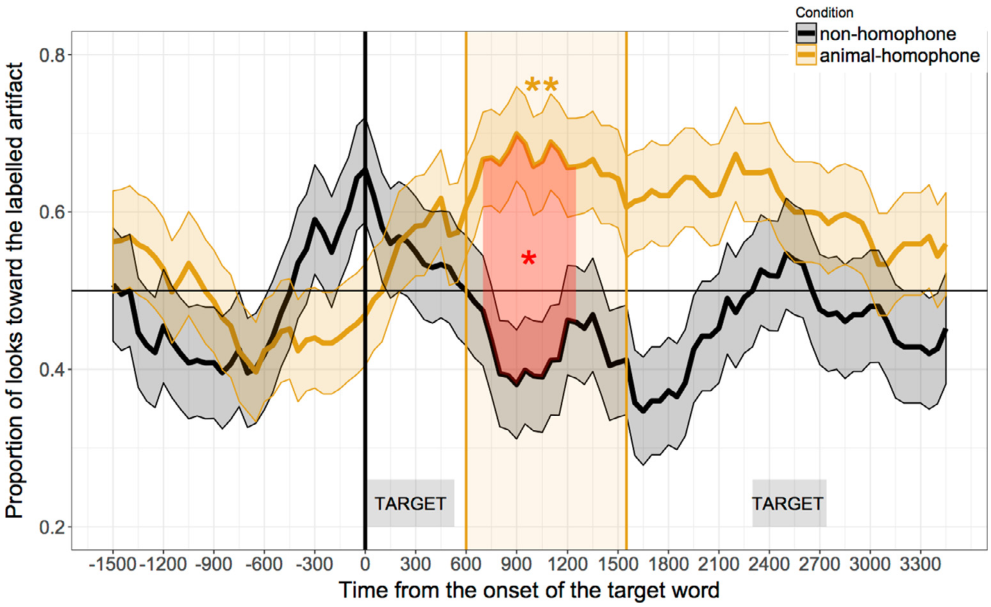

Fig. 6. Proportion of looks towards the animal-homophone referent (an artifact) from the beginning of the first target word (do you see the [target]? where is the [target]? (the gray boxes at the bottom of the plot represent the position, in time, of the two target words within the test sentences during the course of a trial) for the animal-homophone (in gold) and for the non-homophone (in black). The ribbon surrounding each curve represents the standard error of the mean obtained at each time bin for each word condition. Toddlers successfully learnt the animal-homophone: They increased their looks towards the correct picture after target onset significantly above chance (gold shaded area), and this behavior was significantly different when tested on the non-homophone word, in which children tended to switch back to the unnamed object (the red shaded area shows the time-window in which the two conditions are significantly different from one another). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

was used to label an artifact, toddlers increased their looks towards the correct artifact referent (600 ms to 1550 ms; $p = .01$).

Crucially there was a significant difference in performance between the animal-homophone condition and the non-homophone condition (from 700 ms to 1250 ms after target word onset, red shaded area, $p < .05$).[4] This indicates that toddlers treated these two words differently, very much like the artifact-homophone group in Experiment 2 (where artifact-homophones labelled novel animals).

2.3.3. Discussion

Toddlers were able to learn an artifact-homophone when it labelled a novel animal but failed to learn an animal-homophone labeling the same animal (Experiment 2). However when the same animal-homophone was used to label a novel artifact, toddlers

[4] For this experiment, the results yield a significant time window which seems smaller than the effect seen in the figure. This is essentially because the effect is long lasting with a small amplitude, which does not yield a big cluster where the t-values between conditions are greater than 2 (the threshold we set in the cluster analysis). One solution would be to decrease the threshold in order to capture the whole time-window where the two curves are separated. Yet in order to be consistent with our other analyses, we left the threshold as we defined it in the method section.

were able to learn it (Experiment 3). Thus, the results suggest that toddlers have no problem learning a second meaning for a known word, if and only if this second meaning is semantically distinct from the first known meaning.

2.4. Experiment 4 – syntactic distance

In Experiment 4, we investigate whether solely increasing the syntactic distance between the meanings of a pair of homophones may facilitate the acquisition of these meanings. To isolate the effect of syntax from semantic, we focused on grammatical gender. Crucially, grammatical gender is a lexical property, as opposed to a semantic property, in languages where gender categories are not clearly defined in semantic terms (as is the case for most nouns in French, with the exception of nouns designating human beings, and some animals). In gender-marking languages, the gender of the noun determines the form of associated determiners and adjectives. In French, feminine nouns are preceded by a gender-marked definite article “la” or indefinite article “une” and masculine nouns by the gender-marked definite article “le” or the indefinite article “un”, when used in their singular form. Such gender cues have been shown to constrain lexical access in adults (e.g., Dahan, Swingley, Tanenhaus, & Magnuson, 2000; Spinelli & Alario, 2002) but also in young children (Johnson, 2005; Lew-Williams & Fernald, 2007; Van Heugten & Shi, 2009). Interestingly for the current study, adults use such gender-marked contexts to selectively access the meaning of homophones (Spinelli & Alario, 2002). For instance, in French, /sɛl/ means both *a saddle* (feminine) and *salt* (masculine) and each meaning is accessed independently when preceded by a gender-marked article. Gender could thus be used to distinguish between different meanings of the same phonological form, by preventing the activation of lexical candidates that do not belong to the same gender category.

We tested here whether a context marked for grammatical gender can help toddlers to identify a second meaning for a known word when the original and the second meanings are associated with different genders.

In Experiment 4, toddlers were taught that a novel animal label was homophonous with an animal noun they already knew (as in Experiment 2) but this time in a different gender-marked context (**a gender-homophone**, belonging to the same semantic category as the first meaning, but syntactically different; e.g., “une chat”, a $\text{cat}_{\text{feminine}}$, normally masculine in French). If a gender-marking context is sufficient to identify an additional meaning for a known word, then toddlers should recover from their failure to learn an animal-homophone (Experiment 2) and correctly learn it when presented in a different gender-context.

2.4.1. Method

Participants. Sixteen French 20-month-olds took part in this experiment (range = 19;5 months to 20;8 months, mean = 20;3, SD = 0;4, 7 boys). Six additional children were replaced because of fussiness during the experiment resulting in more than 50% of trials with missing eye-tracking data (n = 2), refusal to wear the sticker necessary for eye-tracking (n = 2) and experimental error (n = 2).

Apparatus, procedure and design. Identical to Experiment 2.

Materials. Similar to Experiment 2 expect for the gender of the labels of the animal-homophones.

Target words. The 4 animal-homophones were taught with a different gender (therefore gender-homophones): *une chat, une loup, un poule, un mouche* (cat, wolf, hen and fly) instead of *un chat, un loup, une poule, une mouche*. The average duration of the target words in the test sentences was 530 ms for the gender-homophones.

The 4 non-homophones were identical to those used in Experiment 2.

Measure and Analysis. Identical to Experiment 2.

2.4.2. Results

Fig. 7 shows the proportion of looks towards the referent of the gender-homophone, time-locked to the beginning of the first target word.

Toddlers taught a gender-homophone did not increase their look towards the referent animal at a rate above chance ($p > .2$). Crucially there was no difference between the gender-homophone and the non-homophone condition ($p > .2$) suggesting that they did not learn the gender-homophones.

A replication of this experiment (See Supplemental Material) showed that children patterned the same way than toddlers taught an animal-homophone, providing more evidence that toddlers fail to learn the gender-homophones, in the same way that they failed to learn the animal-homophones (Experiment 2).

2.4.3. Discussion

Experiment 4 attempted to isolate the effect of syntactic context alone in the identification of a novel meaning, by using a gender-marked context. The results show that toddlers failed to take this information into account. Certainly, this does not mean that syntactic context alone is always insufficient to distinguish between the meanings of a pair of homophones, but the present experiment shows that gender may not provide enough evidence for toddlers to conclude that a novel meaning is intended in that situation.

Yet there may be several other interpretations for toddlers' failure. One possibility is that French 20-month-olds do not yet use gender when processing spoken language. However, we believe this possibility to be unlikely: A recent study shows that 18-month-old toddlers can track the statistical dependencies between gender-marked articles and nouns (Van Heugten & Christophe, 2015).

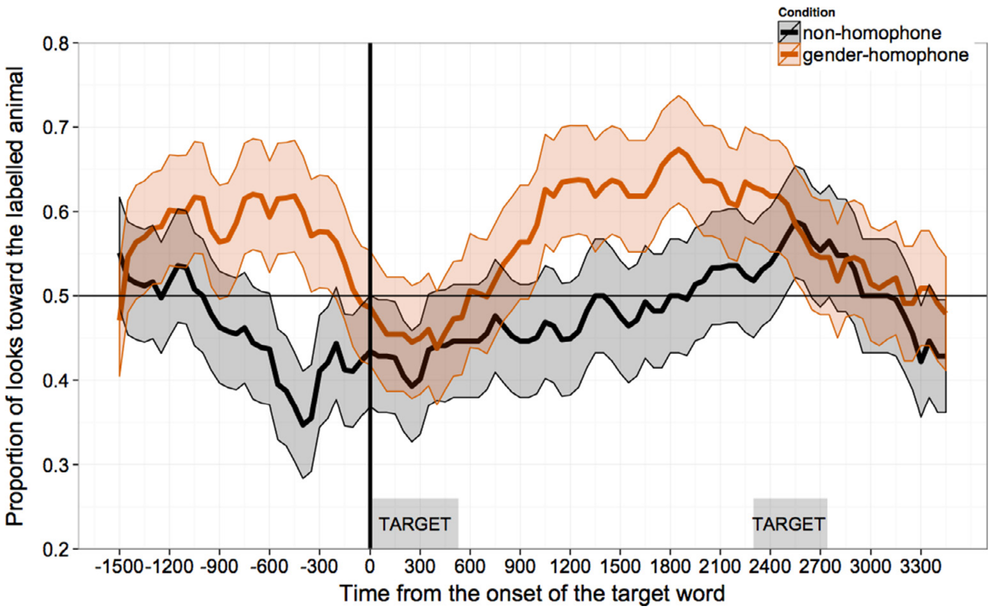

Fig. 7. Proportion of looks towards the target picture, time-locked to the beginning of the first target word (do you see the [target]? where is the [target]? (the gray boxes at the bottom of the plot represent the position, in time, of the two target words within the test sentences during the course of a trial) for the gender-homophone (in orange) and the non-homophone (in black). The ribbon surrounding each curve represents the standard error of the mean obtained at each time bin for each word condition. Toddlers failed to show any recognition of the gender-homophone. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Toddlers preferred to listen to article-noun sequences in which the gender-marked article matched the gender of the noun (e.g., "la$_{fem}$ poussette$_{fem}$", the stroller) than when the gender-marked article mismatched the gender of the noun (e.g.,"le$_{masc}$ poussette$_{fem}$"). This suggests that toddlers of that age may already be sensitive to gender cues when processing speech.

Another possibility is that gender information may not bring convincing evidence that a novel meaning is intended when used to label animals. Indeed, gender marking for an animal-label sometimes (but not always) corresponds to the male and female individuals of the specie: For instance, "un$_{masc}$ chat" /ʃa/ refers to the male cat and "une$_{fem}$ chatte" /ʃat/ to the female cat. Accordingly, if toddlers already know that animals may have different biological genders and understand that female individuals are often preceded by a feminine article (e.g., "la") and male individuals by a masculine article (e.g., "le"), they may have considered that the original meaning was intended when presented with the novel animal, just like in Experiment 2, even though it was preceded by contradictory gender information.[5] Thus, if gender-marking helps to distinguish an additional meaning for a known word form, such information may be available only when labeling non-biological entities. In order to test this hypothesis, future work could use artifact-labels instead of animal-labels.

Finally, toddlers may rely on the phonological form of the test words to determine their gender. Previous work has shown that children rely heavily on phonological cues to noun classes (e.g., Gagliardi & Lidz, 2014; Karmiloff-Smith, 1981) even in the presence of more reliable cues (i.e., semantic cues; see Culbertson, Gagliardi, and Smith (2017) for evidence that children rely more on phonological than semantic cues to gender). In the present experiment although feminine and masculine words can both end in /a/, /u/, /ul/ and /uʃ/ (as in the labels we used), these endings are predominantly restricted to a particular gender (70% of the monosyllabic words ending in /a/ are masculine, 70% ending in /u/ are masculine, 57% ending in /ul/ are feminine and 88% ending in /uʃ/ are feminine; as calculated on a French lexical database, New et al., 2004). Critically, the predominant gender corresponds to the gender of the known label and not to the gender-context we supplied in this experiment. Thus, it remains an open question whether toddlers would be able to learn gender-homophones when the phonological form of the label used is neutral with respect to gender.

At any rate, while toddlers of that age use gender cues when processing speech, such cues may not constitute systematic and reliable evidence to identify that a word form could map onto several meanings, and accordingly toddlers may fail to use it, as we showed here.

2.5. Experiment 5 – neighborhood density

The preceding experiments manipulated the distance (semantic and syntactic; semantic alone; syntactic alone) between a known

[5] However note that this is not always the case 'girafe' is always feminine, even when referring to male giraffes, while 'hippopotame' is always masculine, even when referring to female hippopotamuses.

first meaning and a novel second meaning. Experiment 5 investigated whether the phonological context of the word in the lexicon (whether it belongs to a dense vs. sparse phonological neighborhood) modulates the conditions in which toddlers accept a second meaning for a known word.

Phonological neighborhood density for a word is the number of words that differ from it by one addition, one deletion or one substitution (Luce, 1986). For instance, neighbors of "cat" include words such as "cap", "hat", "fat", "rat", "at", "catch", etc. Some words are said to live in a dense neighborhood when they have many neighbors and some words live in a sparse neighborhood because they have only a few neighbors. Previous work has shown that children consider multiple sources of information when learning words, including the phonological structure of the lexicon already in place. For instance, preschoolers find it easier to learn words that are composed of frequent sound sequences rather than rare sound sequences (e.g., Storkel, 2001) and 20- to 24-month-old toddlers find it easier to learn a new word that has many phonological neighbors in their vocabulary (Newman et al., 2008).[6] One possibility is that children may find it easier to learn a word that re-uses phonological sequences that already exist in the lexicon (Storkel & Maekawa, 2005). Another may be that densely populated areas of the lexicons may be considered as a probable place for a new word to occur (compared to sparse areas), given that so many words already occur there. Since the area is densely packed, it also becomes reasonably probable that a novel word form should fall in a place that is already occupied by a known word, and thus leads to homophony. In contrast, a novel word appearing in a sparse neighborhood is more unlikely to start with, and the probability that in this empty space, it would fall exactly in the same spot as the only existing word, is suspiciously low. Toddlers may thus resist learning a novel meaning for a known word living in a sparse phonological neighborhood. Following this idea, in the preceding experiments, we purposefully selected known words with a high phonological neighborhood density in toddlers' lexicon, to maximize our chances of finding learning in some situations. On average the animal-homophones had a phonological neighborhood density of 3.1, and the artifact-homophones of 3.8,[7], which is rather high considering that an average French 20-month-old toddler comprehends about 200 words (according to measures using the French CDI in previous experiments (Kern, 2007).

In Experiment 5 we directly manipulated neighborhood density to test whether it modulates the learning effect observed in Experiment 2. Toddlers were taught that a novel animal label was homophonous with an artifact noun they already knew (as in Experiment 2) but this time the label was from a sparse phonological neighborhood in toddlers' lexicon (a **sparse artifact-homophone**: belonging to a different semantic category than the original meaning, and whose word form has a low phonological neighborhood density; e.g., "un livre", *a book*). If phonological neighborhood density helped toddlers to learn the dense artifact-homophones in Experiment 2, then toddlers may fail to learn the sparse artifact-homophones. On the contrary, if toddlers are able to learn the sparse artifact-homophones from the present experiment as well as the dense artifact-homophones from Experiment 2, this would suggest that increasing the semantic distance between the meanings of a pair of homophones is enough to learn these meanings, independently of the phonological neighborhood density of the word form.

2.5.1. Method

Participants. Sixteen French 20-month-olds took part in this experiment (range = 19;1 months to 21 months, mean = 20, SD = 0;4, 4 boys). Seven additional children were replaced because of fussiness during the experiment resulting in more than 50% of trials with missing eye-tracking data (n = 4), refusal to wear the sticker necessary for eye-tracking (n = 1), eye-problems on the day of the experiment (n = 1) and experimental error (n = 1).

Apparatus, procedure and design. Identical to Experiment 2.

Material. Similar to Experiment 2 except for the set of sparse artifact-homophones used in the teaching phase.

Target words. The 4 sparse artifact-homophones were also all monosyllabic words: livre, fleur, fraise, sieste (/livʁ/, /flœʁ/, /fʁEz/, /siɛst/) meaning: *book, flower, strawberry* and *nap.*[8] These words had an average phonological neighborhood density of 0 in children's lexicon (according to the parental report) and an average frequency count of 207 in a corpus of child directed speech (the Lyon corpus, Demuth & Tremblay, 2008). The average duration of the sparse-homophones in the test sentences was 726 ms.

Measure and Analysis. Identical to Experiment 2.

2.5.2. Results

Fig. 8 shows the proportion of looks towards the referent of the sparse-homophone time-locked to the beginning of the first target word.

Toddlers taught a sparse artifact-homophone behaved like toddlers taught a dense artifact-homophone: They looked to the correct referent at a rate above chance (in light blue; from 1550 ms until 3500 ms; $p < .001$) and there was a significant difference between the sparse artifact-homophone condition and the non-homophone (in dark blue; from 1750 ms until 2600 ms; $p < .001$), suggesting

[6] This may appear surprising, since we and others (Dautriche et al., 2015; Swingley & Aslin, 2007) have found that 18-month-olds have a hard time learning phonological neighbors. Yet as shown by Newman et al. (2008), what actually seems to matter is the number of phonological neighbors: while one phonological neighbor impairs learning (Swingley & Aslin, 2007; Dautriche et al., 2015), many neighbors may provide a small benefit (Newman et al., 2008).

[7] To calculate neighborhood density for our test words, we gave a questionnaire to parents including all existing neighbors of the test words in the adult lexicon and ask parents to indicate which ones their toddler understands. The average neighborhood density obtained corroborated our prediction while choosing the stimuli: For this group of toddlers at least, both the animal-homophone labels and the artifact-homophone labels had a comparable neighborhood density.

[8] As the reader may have noticed, "nap" is not an artifact. Because the set of monosyllabic and dense words that toddlers of that age know is quite small, we had to broaden our search to the set of non-animal labels.

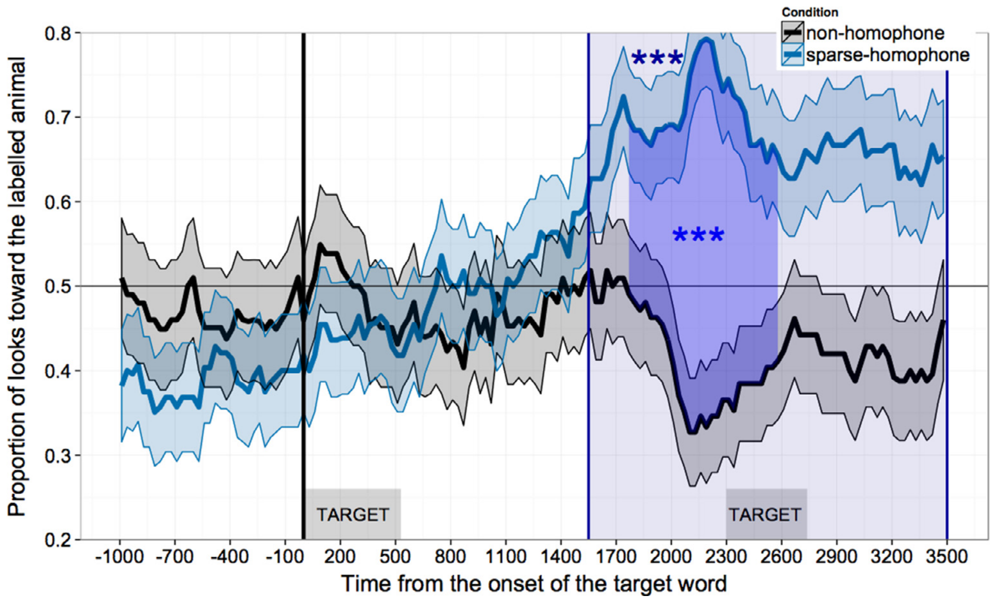

Fig. 8. Proportion of looks towards the sparse artifact-homophone referent (an animal), time-locked to the beginning of the first target word (do you see the [target]? where is the [target]? (the gray boxes at the bottom of the plot represent the position, in time, of the two target words within the test sentences during the course of a trial) for the sparse- homophone (in blue) and for the non-homophone (in black). The ribbon surrounding each curve represents the standard error of the mean obtained at each time bin for each word condition. Toddlers had no problem to learn a sparse-homophone, they looked to the target animal significantly above chance in the sparse artifact-homophone condition (light blue shaded area) and this behavior was significantly different from the non-homophone condition (dark blue shaded area). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

that they learnt the sparse artifact-homophone and treated it differently from the non-homophone.

2.5.3. Discussion

Toddlers successfully learnt a sparse artifact-homophone, and behaved in the same way as if they learnt a dense artifact-homophone (Experiment 2). This suggests that phonological neighborhood density does not exert a major influence on homophone learning, at least in the present experimental conditions.

Yet this result may be surprising given the past literature suggesting a learning advantage for words coming from dense neighborhoods in lexical acquisition (Coady & Aslin, 2003; Newman et al., 2008; Storkel & Hoover, 2011; Storkel & Lee, 2011). One possibility is that such a learning advantage is at play when learning a new word form as opposed to a new meaning as previous studies on the subject do not distinguish between these two aspects of word learning. Another possibility is that, in this study, the difference between the dense artifact-homophones (having between 2 and 5 neighbors) and the sparse artifact-homophones (with 0 neighbors) may not be sufficient to observe an effect. However, increasing this difference is challenging, as the lexicon of 20-month-olds is still very sparse at this young age. Finally a last possibility is that this experimental paradigm is not sensitive enough to observe a difference between these two conditions. Indeed, if there is a small disadvantage for sparse labels to be homophones, it might not be observed here because learning is already at ceiling for sparse homophones. To test that hypothesis, one could make the task harder, for instance by removing some of the information in the learning videos, and see whether a difference between sparse and dense labels could be obtained in these conditions.

At any rate, this suggests that the influence of neighborhood density, if any, is smaller than the influence of context (semantic and syntactic) on homophone learning, at least in these experimental conditions.

2.6. Interim discussion

Taken as a whole, this set of studies shows that toddlers have no problem learning homophones, as long as the meanings associated with the same word form appear in contexts that are sufficiently distinct for children.

To our knowledge, this is the first study reporting homophone learning in such a young age group. This result may be surprising given the consistent failure of older children (preschoolers) to learn a second meaning for a known word (Casenhiser, 2005; Doherty, 2004; Mazzocco, 1997; but see Dautriche, Chemla, & Christophe, 2016, when the two meanings are learnt simultaneously). However, previous research on homophone learning differs from ours in two important aspects. First, previous studies typically did not manipulate whether the learning situation proposed to children could plausibly lead them to conclude that an additional meaning was likely for that known word (although Casenhiser, 2005, report children's success in learning a second meaning for a known word when this known word is placed in a different syntactic frame). Second, they also relied on rather poor learning situations, in which the word was used only a limited number of times (once or twice) in stories that failed to provide sufficient evidence to constrain a potential novel meaning for the word. In particular, it may not be sufficient for children to know that the known word form "door" refers to an unfamiliar animal (e.g., a tapir), as the referent of the word is still ambiguous (e.g., Quine, 1960). Children may need additional evidence about what properties are associated with the new meaning of "door" (e.g., living in the jungle, eating berries) to narrow down its meaning. In our experiments, we use a word learning task which has the advantage of presenting novel words in a richer context than in previous studies on preschoolers: the homophonous label is repetitively used with a supporting visual context (i.e., the novel referent), social support (i.e., the speaker looks back and forth between the child and the named visual referent, making clear what she is talking about, and providing cues that she is teaching something to the child, i.e. Csibra & Gergely, 2009) and supplemented with information about its novel meaning (e.g., "Un manger, ça a des grandes oreillles" *Eats have big ears*). This rich learning situation gave us the possibility to manipulate various aspects of the to-be-learnt meaning, to figure out what properties of the novel word enabled toddlers to learn it (its syntactic context, the semantic proximity to other words in the lexicon, and the position of its label in phonological space) in this maximally supportive learning context.

While our set of studies tackles several of the questions and experimental difficulties seen in previous work of homophone learning, it also opens an avenue for future research. Due to the limitations pertaining to testing young children, we focused on a limited set of contrasts that may help to distinguish between two meanings of a pair of homophones. To be more explicit, Experiment 1 investigated only the noun/verb distinction, Experiments 2 and 3 focused on the animal/artifact semantic distinction, Experiment 4 focused on a single morphosyntactic property: gender and finally Experiment 5 on a single phonological property of words, their neighborhood density. An interesting question is whether these results can be found using a broader set of contrasts (syntactic and/or semantic), in other languages or even by looking at other contrasts not investigated here. We believe that the simple paradigm we developed here may be used to look into these questions.

To sum up, our experimental results suggest that the process of creating a novel lexical entry is mediated by multiple sources of information coming from the lexicon (e.g., lexical-semantic relationships) and the parsing system (e.g., expecting a noun or a verb in a given linguistic context). As a result, homophones are less challenging for children than previous studies suggested, at least whenever each member of the pair appears in distinct syntactic or semantic contexts. One important question is whether the structure of the lexicon reflects these constrains on learning: Is it the case that members of a homophone pair are more distant from one another than would be expected by chance alone? Such a result would suggest that the lexicon might be shaped by learning constraints.

3. Corpus analysis studies

Learning homophones may be difficult even for preschoolers (Casenhiser, 2005; Doherty, 2004; Mazzocco, 1997), yet this difficulty is reduced when the two meanings of a pair of homophones have different syntactic categories or cover distinct concepts. One interesting question is thus whether these learnability advantages translate into the overall structure of the lexicon of natural languages: Are there more homophones from different syntactic categories than from the same syntactic category? Similarly, are members of a homophone pair more likely to be semantically distant? Interestingly, the present results also suggest that grammatical gender does not help toddlers to identify whether a given word form maps onto several meanings. Following the same idea, this suggests that grammatical gender might not exert a major influence on the organization of homophony within the lexicon. The case of neighborhood density is less clear, as its potential impact on learnability may be hidden by a ceiling effect due to the richness of the teaching context we used, see the discussion of Experiment 5.

To investigate these questions, we extracted the pairs of homophones in the lexicon of 4 languages (Dutch, English, French and German) and computed several measures looking at their syntactic category, their semantic distance, their gender, and the neighborhood density of the phonological form. We then compared these measures to random baselines that simulate how homophone pairs should be distributed under random conditions, if there was no learning pressure exerted on the set of homophones in the lexicon.

3.1. Method

Lexicons. We used the phonemic lexicons of 4 languages: Dutch, English, German (extracted from CELEX, Baayen, Piepenbrock, & Gulikers, 1995) and French (extracted from Lexique, New, Pallier, Brysbaert, & Ferrand, 2004). For each language, we removed the stress marks of words[9] and excluded hyphenated words (words containing a space character or a hyphen) as well as words which had no grammatical tags. This resulted in a lexicon of 341,901 words for Dutch, 118,465 words for English, 137,650 words for French and 232,728 words for German.

Homophone identification. To identify homophone pairs in each lexicon, we listed all the pairs of words that shared the same phonological form but had different lemmas according to their lemma code in CELEX and Lexique. This procedure eliminated homophones coming from the same root but instantiated by different categories (e.g., to fight/a fight) or where one of the forms has a silent morphological marker (e.g., chien/chiens *dog*/*dogs*, which are pronounced in the same way in French). We kept only a single homophone pair for a given lemma pair (e.g., we eliminated flowers/flours if flower/flour was already present in our list). This

[9] Because the stress marks were removed, noun-verbs pairs of words which are distinctive by the position of stress were counted as « pure » homophones (e.g., « desert » in English).

resulted in 12,748 homophone pairs for Dutch, 10,652 for English, 8135 for French and 1657 for German.

Measures. *Syntactic category.* We used the Part Of Speech (POS) tags in CELEX for Dutch, English and German and in Lexique for French, to count the number of homophones within the same syntactic category (e.g., animal-“bat”/baseball-“bat”) and the number of homophones across different categories (e.g., a park/to park).

Semantic Similarity. To derive a measure of similarity between words, we used Latent Semantic Analysis (LSA; Landauer & Dumais, 1997), a class of distributional semantic models that builds on the idea that similar word meanings occur in similar contexts. For instance, in a sentence context such as “we are eating X for lunch” X is likely to be some edible object but not a vehicle or a piece of furniture. We applied LSA on Wikipedia for each language using the Gensim package (Řehůřek, Sojka, & others, 2010) in python. Each (orthographic) word of the Wikipedia corpus was modeled as a vector in a multidimensional space (of dimension $d = 400$ following previous studies, Řehůřek et al., 2010) obtained by counting the number of occurrences of all the words found in every chunk of text (a document) where this word appears. All words that appeared in less than 5 documents and in more than 10% of the documents were discarded. This was to ensure the quality of the semantic representation, since low and high frequency words are poor predictors of semantic content. After applying these constraints, we kept a vocabulary of the 100,000 most frequent words in order for the model to be tractable. Note that these constraints may discard homophones if they are too frequent or too infrequent or are not in the 100,000 most frequent vocabulary after applying these constraints. The semantic similarity between two words is the cosine of the angle between the two word-vectors, and ranges between -1 and 1. A value close to 1 or -1 indicates that two words are close in meaning (a value close to 1 means that the two word-vectors have the same orientation much like synonyms and a value close to -1 are two-word vectors that have opposite direction much like antonyms), whereas values close to 0 indicate that the meanings are not related (the two vectors are orthogonal). Because we are interested in the angle between word-vectors and not their orientation, we took the absolute value of the cosine similarity (however taking into account both negative and positive values does not change the pattern of results presented below).

We focused on same-category homophones that have a different spelling. For instance, we excluded “bat” whose two meanings are written in the same way, and would lead to the same semantic representation in the semantic space we use, which is based on orthographic words (LSA, as explained above). Native speakers of these languages excluded spelling variants (e.g., analyse/analyze) which would be counted as homophones and would by definition be semantically very close.[10] Finally, due to the restriction on lower and higher frequency words on the semantic model (see above), this led to 121 pairs in Dutch, 268 in English, 241 in French and only 40 in German (mostly due to the presence of low frequency words in the different-orthography homophone pairs we could extract).

On that subset of words, we computed two measures based on cosine similarity: first, we averaged the cosine similarity of all the pairs of homophones having different spellings (see in the supplemental material D., an analysis using the English WordNet for the same measure); second, we grouped the homophones in 5 bins of semantic relatedness (from not related at all to highly semantically related), and computed the absolute number of homophones within these similarity bins.

Gender. We focused on noun-noun homophones leading to 1478 pairs in Dutch, 1338 in French and 220 in German (the three languages which instantiate grammatical gender). We used the gender information tags provided in CELEX for Dutch and German and in Lexique for French to count the number of noun-noun homophones within the same gender (e.g., “avocat”, meaning avocado$_{masc}$ or lawyer$_{masc}$) and across different genders (e.g., “mur”/”mûre”, wall$_{masc}$/blackberry$_{fem}$). Note that there are 3 grammatical genders in Dutch and German (feminine, masculine, neutral) and 2 in French (feminine, masculine).

Neighborhood density. We computed the average number of neighbors for the homophonic forms in the lexicon of the 4 languages under study.

Random baselines. Each random baseline was repeated 1000 times in order to obtain a chance distribution for each measure.

Syntactic category. For each language, we shuffled the syntactic categories within each word length to create a suitable random baseline to evaluate the number of pairs of homophones that fell across syntactic categories.

Semantic similarity. For each language, we randomly shuffled the LSA vectors within all words of the same length (for all the lexicon) to create a suitable baseline to evaluate the semantic similarity between the meanings of homophone pairs.

Gender. For each language, we extracted the subset of nouns from the lexicon and shuffled their grammatical gender within each word length.

Neighborhood density. For each homophonic form, we picked another form of the same length in the lexicon, and computed the neighborhood density of the chosen word form.

Statistical analysis. Essentially this method follows the logic of hypothesis testing: We compute a test measure (m) and compare it to a distribution of this measure under the hypothesis that the pairs of homophones in the lexicon are distributed randomly with respect to this property. We then evaluate whether the value for the real lexicon falls outside of the distribution of the random baselines. Because the null distribution of our measures is close to normal (see Figs. 9–13), we can compute a mean (μ) and a standard deviation (σ), to calculate a z-score ($(m - \mu)/\sigma$), its 95% confidence interval, and its associated p-value.

3.2. Results

3.2.1. Across-category homophones in the lexicon

We first considered the proportion of homophone pairs that are distributed across syntactic categories. If there are more

[10] The manual exclusion was done only for this analysis as the number of pairs of homophones was tractable. Note that because the spelling variants are from the same category this does not bias the proportion of across-syntactic-category homophones.

homophones across syntactic categories than expected by chance, the proportion of across-category homophones should be greater in the lexicons than in the random baselines.

Fig. 9 shows how the random baselines (the histograms) compare to the lexicons (the red dots). First, we can notice that the proportion of across-category homophones (ranging from 0.6 to 0.9 across the 4 languages) is greater than the proportion of same-category homophones (the complementary proportion). Second, and this is crucial here, all histograms fall to the left of the red dot, which means that all lexicons have more across-category homophones than expected by chance (all $p_s < .001$). This is consistent with Ke (2006) who found the same results for homophones of Dutch, English and German. This suggests that there is a pressure for homophones to be distributed across syntactic categories rather than within the same syntactic category.

3.2.2. Semantically unrelated homophones in the lexicon

Using the semantic similarity measure derived from the LSA analysis, we first looked at the average semantic similarity of all pairs of homophones from the same grammatical category (and whose members have a different orthography, see the Method section).

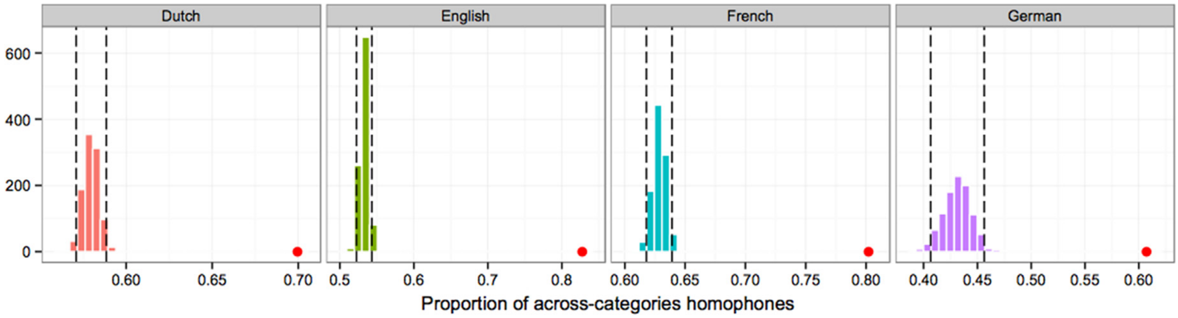

Fig. 9. These histograms show the distribution of the proportion of across-category homophones compared to the real lexicon (the red dot). The dotted lines represent 95% confidence intervals derived from the distribution of random baselines. All 4 languages have significantly more across-category homophones than expected by chance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

As seen in Fig. 10, the average semantic similarity between meanings of a homophone pair does not differ from chance across the 4 languages (although this was marginal for English; $p = .052$, with a tendency for homophones to be less similar in the real lexicon than in the random baselines, which is the direction expected from the experimental results).

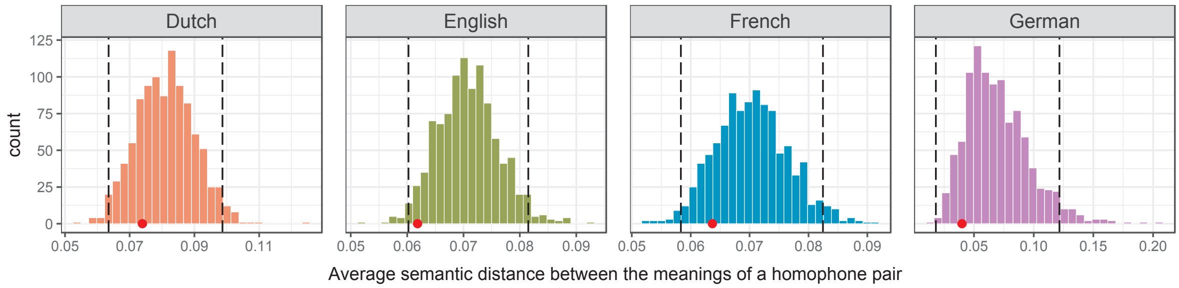

Fig. 10. These histograms show the distribution of the average semantic distance between members of a pair of homophones in our random baselines compared to the real lexicon (the red dot). The dotted lines represent 95% confidence intervals derived from the distribution of random baselines. The members of the pairs of homophones in these 4 lexicons are not more semantically distant than expected by chance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

This result, however, should not be surprising. Recall that the chance level was constructed by randomly assigning a meaning to each word involved in a homophone pair. Yet, how likely is it for two randomly chosen meanings in the lexicon to be similar? The intuition is that it is pretty low.[11]. As a result, the chance level *already* measures the absence of similarity between two meanings. Thus if homophones are indeed semantically distinct, then it is no surprise that their average semantic distance does not differ from such a chance level.

Since the average semantic similarity across all pairs of homophones (in the lexicon) is arguably not the best way to detect a potential lack of semantic similarity within homophones, in a second analysis, we looked at the distribution of homophone pairs

[11] To get an idea of the semantic similarity of words on the cosine scale, here are a few examples of pairs of words for the English lexicon with their cosine in parenthesis: prize/cigar (0.003); lethal/static (0.11); classify/simplify (0.29); headline/telegram (0.41); electricity/consumption (0.55); hormone/vitamin (0.61); yellow/purple (0.72); cookery/vinegar (0.84); reactor/fission (0.96).

compared to random pairs of words on the cosine scale. We divided the range of cosine values (from 0 to 1, where values close to 0 indicate that the meanings are not related whereas values close to 1 indicate that two words are close in meaning) into 5 uniformly spaced bins and counted the number of homophone pairs falling into each bin for the real lexicon and for each of the random baselines. If there are less semantically related words among homophones than among words in the lexicon, homophones should be absent (or less numerous than in the random baselines) in the cosine bins that capture the most similar words. Note that here, as before, we compare the same number of words pairs in the real lexicon (pairs of homophones) and in the simulated baselines (pairs of random words), but we now look at their distribution along the cosine scale.

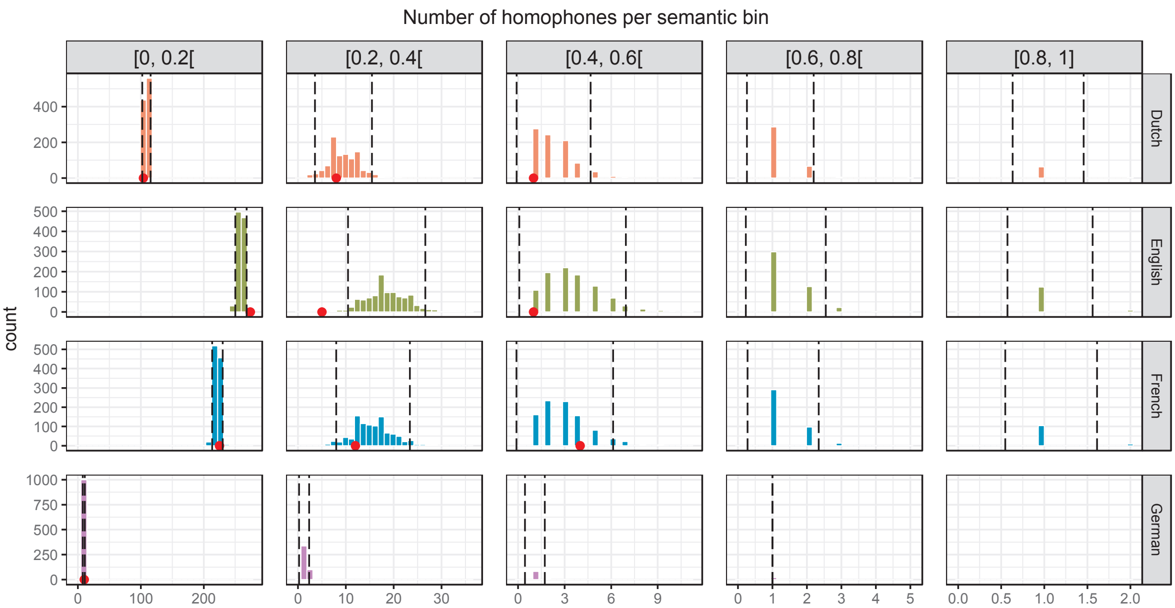

Fig. 11. Number of homophone pairs per semantic bin. The histograms show the number of pairs of homophones found in the random baselines for each 0.2 bin of cosine value compared to the number of pairs of homophones in the lexicons (red dots) in the same bin. There is no pair of homophone in the bins that capture the most similar words (values close to 1, rightmost graphs), despite the fact that the random baselines span the full range of cosine values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

This is indeed what we observe: There is no homophone in the two last bins on the right (cosine values from 0.6 to 1) in Dutch, English and French, and no homophones in the last 4 bins (0.2–0.8) for German. Certainly, not all random lexicons have homophone pairs with values in these bins: For instance, in English about 125 of the 1000 random baselines have one pair of homophones with a cosine value greater than 0.8 and about 450 have more than one pair of homophones with a cosine ranging between 0.6 and 0.8. This means that there is between 40% and 60% of random lexicons[12] that have values in these two last bins (for English, but the same can be established for the other languages), meaning that the chance of *not* getting a homophone within these bins is also situated between 40% and 60%. While this does not lead to statistical significance, the pattern is consistent across the 4 languages, and suggests that semantically highly similar homophones seem to be dispreferred in languages.

In sum, our results suggest that semantically highly similar homophones are dispreferred in the lexicons of natural languages. A similar analysis was conducted in English, with a much larger set of homophones, using WordNet distances (see the Method for the constraints imposed by the LSA analysis), and its results converge with the present findings (see Fig. S5 in the supplemental material). Yet because most pairs of words are dissimilar in the lexicon (i.e., two randomly-picked meanings are likely to be dissimilar), we do not observe differences in the average semantic similarity between homophones and randomly selected meanings (Fig. 10), the difference only becomes apparent when one looks at the distribution of cosine values (Fig. 11).

3.2.3. Across-gender homophones in the lexicon

We considered the proportion of noun-noun homophone pairs that are distributed across different genders. Because English is not a gender-marked language, this analysis focuses on Dutch, German and French.

As shown in Fig. 12, the proportion of across-gender homophones is lower than chance (Dutch and French: $p_s < .001$) or undistinguishable from chance (German): this is the opposite result of what we would expect if same-gender homophones were dispreferred because they were harder to learn. This suggests that there is no pressure for distributing noun-noun homophones across

[12] 40% if the 125 lexicons that have value in the cosine bin 0.8–1 are a subset of the 450 that have value in the 0.6–0.8 bin and 60% if the 125 lexicons that have value in the cosine bin 0.8–1 are a different set of lexicons from the 450 lexicons that have value in the 0.6–0.8 bin.

grammatical genders. Yet it may also be the case that a pressure to attribute different genders to the different meanings of a homophone is counteracted by other pressures, such as phonological correlates of gender, that would make it difficult to get different-gender homophone pairs – but would make it easier to learn and remember gender itself (e.g., Cassidy, Kelly & Sharoni, 1999; Kelly, 1992).

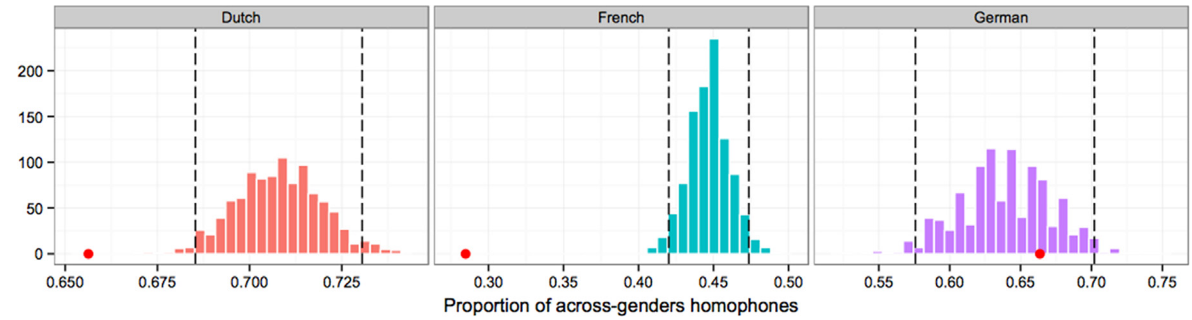

Fig. 12. These histograms show the distribution of the proportion of across-gender homophones compared to the real lexicon (the red dot). The dotted lines represent 95% confidence intervals derived from the distribution of random baselines. Lexicons do not have more different-gender homophones than expected by chance, if anything they have less of them (in Dutch and French). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2.4. Neighborhood density

Finally, we looked at the average neighborhood density of the homophones in the four languages under study. As we can see in Fig. 13, homophones in these four languages have, on average, more phonological neighbors than expected by chance (all $p_s < .001$).

This was expected given that word forms that have lower phonotactic suprisal are more likely to map to several meanings (Piantadosi et al., 2012) and that lower phonotactic surprisal is correlated with higher neighborhood density (Mahowald, Dautriche, Gibson & Piantadosi, in review). This, however, highlights a potential confound, as the result we observe might not be due to neighborhood density but to phonotactic surprisal. We conducted another analysis controlling for phonotactic suprisal. To do this, we selected words constituting the random lexicon not only to match the length of homophones but also to match their phonotactic surprisal as closely as possible.[13] However this did not change the main result: Homophonous word forms still have a higher neighborhood density than expected by chance, even when controlling for phonotactic surprisal (all $p_s < .001$). This suggests that homophony is more likely to occur in densely packed regions of the lexicon. While our results with toddlers do not highlight an advantage to learn additional meanings for word forms coming from dense regions of the lexicons, we cannot conclude that the experimental and the quantitative analyses presented here are in contradiction. A potential advantage for homophones from dense neighborhoods, if present, may be hidden by a ceiling effect triggered by our experimental procedure, as we will outline in the following discussion.

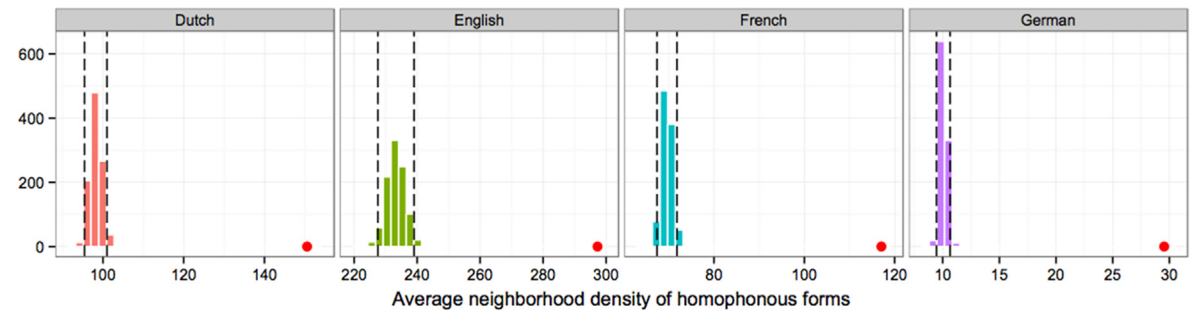

Fig. 13. Average neighborhood density among all homophonous forms (red dots) compared to the chance level (histograms). In all 4 languages, the average neighborhood density of homophones is higher than what would be expected by chance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.3. Discussion

Homophony is a by-product of language contact and sound change (e.g., Kaplan, 2011; Wedel et al., 2013). Yet the distribution of homophones in the lexicon is not random and mainly reflects ease-of-learning by toddlers, as observed in the experiments reported

[13] We trained a trigram model on phones (with a Laplace smoothing of 0.01 and with Katz backoff in order to account for unseen but possible sound sequences) on each lexicon and used the resulting model to find the probability of each word under the model (a proxy for phonotactic surprisal).

above, with easy-to-learn homophones being over-represented in the lexicon: (1) there are more across-category homophones in the lexicon than same-category homophones, and it is also the case that toddlers learn homophones easily when they span different syntactic categories (Experiment 1); (2) Homophones are likely to be semantically more distinct, and toddlers find it easier to learn homophones when they cover distinct concepts (Experiment 2); (3) there are no more across-gender homophones than expected by chance, and children seem to not rely on gender to learn homophones (though see the potential limitations of these results in the Discussion of Experiment 3); However, (4) although homophonous word forms have a denser phonological neighborhood than non-homophonous word forms, Experiment 4 did not reveal better learning for dense over sparse homophones; as noted above, this might be due to a ceiling effect (since learning is perfect in both conditions), and the effect of phonological neighborhood on homophone learning (if it exists) may be too subtle to disrupt learning in the experimental paradigm we developed for children (see discussion of Experiment 4).

This trend was observed for all four languages under study (Dutch, English, French and German) suggesting that these results are robust enough. While it would be tempting to conclude for a global cross-linguistic pattern of homophony favoring disambiguation in context, we note that we only considered four languages in these analyses, of which three are from the same language family. There may be also other features in languages – not considered here – that help to distinguish words in context and thus may facilitate homophone learning and/or serve as a dimension onto which homophones are organized to be maximally distinct. Future work extending these results to more language families and studying different features of languages would be necessary to conclude that what we observed here is a universal feature of lexical ambiguities in language, which would follow if this pattern stems from cognitive pressures.

Quantifying lexical properties is subject to a number of constraints imposed by the lexical databases we used. In particular, all these analyses depend on the coding scheme used for lemmas: Which word form counts as a lemma and which word forms are derived from it. This was essential to spot homophones which, by definition, belong to different lemmas. Yet it certainly misses homophones on the way. For instance, “to run” and “a run” are coming from the same lemma according to CELEX. Yet “a run” could mean, inter alia, a score in baseball while “to run” could mean to manage a place. This is a case where these distinct meanings could well be considered as homophones and not as derived meanings of the same base (polysemes). In addition, spelling variants (e.g., analyze/analyse) are counted as separate lemmas in both CELEX and Lexique. While we excluded such pairs of words in the semantic analysis (see the method section), the presence of these words may have impacted the proportion of homophones in the other analyses. However, even though our method for identifying homophones was not perfect (missing some, and detecting a few spurious ones), our conclusions are always based on the comparison between the kinds of homophones attested, and what would be expected by chance (based on the same set of homophones). Therefore, any imperfection in the lemma coding may not have induced spurious results.

To sum up, the present results show that the distribution of homophones in the lexicon is not random and that there are some correspondences between what makes homophones easy to learn and how they are organized in the lexicon. We discuss this further in the General Discussion.

4. General discussion

An important part of the word learning process requires children to identify what counts as a novel word and what does not. This is especially a challenge when learning homophones, where the same phonological form is used to refer to several distinct meanings. Here, we investigated different sources of information that may help children to identify when it is appropriate to assign a novel meaning for a known word form. Specifically we manipulated (1) the syntactic and the semantic distance between the novel word and its familiar homophone (Experiments 1–4) and (2) the position of the word form in the phonological network of the mental lexicon (Experiment 5).

Experiment 1 showed that toddlers had no problem learning homophones when their meanings are realized in different syntactic categories: “an eat” was a good label for a novel animal, despite children knowing the meaning of the verb “to eat”. Yet, this does not tell us whether the syntactic distinction (noun/verb) or the semantic distinction (object/action) between the two meanings was what allowed toddlers to learn a novel homophone. Experiments 2 and 3 showed that a semantic distinction between the two meanings of a pair of homophones was sufficient to trigger learning: Toddlers learnt easily that “a glass” could also label a novel animal but failed in a condition where the novel animal was labelled “a cat”. However, Experiment 4 failed to show that the syntactic context alone, when different meanings of a homophone are cued by different genders, is sufficient to learn a second meaning for a known noun: Toddlers failed to learn that a novel animal could be called “une_{fem} $\text{chat}_{\text{masc}}$”/ *a cat*, despite the presence of the article “une_{fem}”. Finally, Experiment 5 suggests that if the phonological density of the word form of the homophone has an influence on establishing that this word form maps onto several meanings, it is not sufficient to make toddlers fail to learn sparse artifact-homophones in our experimental design.

The experimental results gathered here suggest that the learning system of young children is equipped with constraints and mechanisms that allow them to successfully learn homophones, as long as they can be distinguished by some contextual elements that children can capitalize on. Thus, children can deal with word form ambiguity as long as distinctiveness is maximized along other dimensions that are relevant for them. If learnability influences the lexicon over the course of language change, then we expect this constraint on early lexical acquisition to have a long-lasting impact on the overall structure of the lexicon. In other words, when homophones occur, are they preferentially distributed across syntactic or semantic categories to improve their learnability?

To answer these questions we looked at the distribution of homophones in the lexicon of four languages. To establish whether the distribution of homophones along these dimensions are different from what would be expected by chance, we compared it to a baseline that represents what the lexicon would look like if homophones were randomly present in the language (see Dautriche, Mahowald, Gibson, Christophe and Piantadosi, 2017, for a similar methodology). Our results show that homophones appear across distinct syntactic categories in the lexicon at a rate greater than expected by chance, and that they are more likely to be semantically distinct. In contrast, the gender distinction does not seem to be a dimension that matters to distinguish homophones. These results are consistent with the hypothesis that languages select homophones that are learnable by children. Since they were observed in four languages, Dutch, English, German and French, there may be a trend across languages to select learnable homophones, although more evidence should be gathered to establish whether this constitutes a universal pattern in the distribution of homophony, across the world's languages.

Taken together our results show that there are some correspondences between what makes homophones easy to learn and how they are organized in the lexicon (see the discussion of Experiment 5 for the case of neighborhood density). Certainly this does not imply that learnability constraints translate directly into lexical structure. The distribution of homophones in the lexicon is also compatible with a pressure for communication: If homophones can be easily distinguished in context, it is also more likely that their meanings will be transmitted accurately. For instance, it seems unlikely that one could be confused about the meaning of the word "bat" in a sentence such as "Bats are present throughout most of the world, performing vital ecological roles of pollinating flowers and dispersing fruit seeds." (extracted from Wikipedia), because the meaning baseball-bat is unlikely in this context. Our results cannot distinguish between the influence of learning and of communication in the distribution of homophones in the lexicon, but bring evidence that the homophones that are currently in the language display properties that make them learnable (and easy to access during comprehension). This suggests, tentatively, that homophones that did not display these properties may have been eliminated from the lexicon across language evolution. Interestingly, Bloomfield (1933) reports that in a dialect of Southwestern France, when the Latin forms "gallus"/*rooster* and "cattus"/*cat* were in danger of merging into one form, "gat", another novel word acquired the meaning *rooster*, suggesting that the use of the same label for *cat* and *rooster* was unwanted and caused speakers to remap a new form onto one of these meanings. This illustrates that pairs of homophones that belong to the same semantic field tend to be eliminated during the course of language evolution. Relatedly, Wedel, Jackson, and Kaplan (2013) report that a phoneme pair (e.g., b/p) is less likely to merge over the course of language change if it distinguishes a large number of minimal pairs from the same syntactic category (i.e., pairs of words that differ by only one sound such as "cab" and "cap"). This implies that the homophones resulting from phonological mergers are preferably distributed across syntactic categories and thus, preserve their distinctiveness.

Lexical ambiguity (and ambiguity in language in general) has been thought to be a great flaw of the linguistic system (e.g., Chomsky, 2002). Our result, together with previous research (e.g., Piantadosi et al., 2012; Wasow, Perfors, & Beaver, 2005), points towards a functional explanation for the presence of homophones in the lexicon. Let us illustrate this with an example. Imagine for instance a language with a limited phone inventory {b, a} that forces its words to be maximally distinctive. Intuitively one may start to form words using the shortest forms possible, such that "ba". Yet because this language maximizes the distinctiveness of its words, such a language needs longer words to express more meanings, for instance: "baba", "bababa" and so on. It is easy to see that a language with a hard constraint for distinctiveness will have many words (as one word can have only one meaning) and therefore will need to rely on long, and complex, words. This language will be difficult to memorize, hard to produce and to process. Relaxing a hard constraint on distinctiveness, and thus allowing homophones in languages, allows the language to use short and easy word forms to convey multiple meanings (Piantadosi et al., 2012). Some have proposed that such a compressible lexicon may be easier to learn for children (Storkel & Maekawa, 2005; Storkel, Maekawa, & Aschenbrenner, 2013) as it minimizes the amount of new phonological information that must be represented in the lexicon. For instance, to learn a novel word such as "blick", children need to create a novel phonological representation /blIk/ that needs to be associated to a novel semantic representation. Learning several meanings for the same phonological form may be more efficient because children only need to learn a novel semantic representation that they can associate with an already existing phonological representation. However, even if a compressible lexicon may convey some learning advantages, we are still left with the main problem posed by homophones: How is it possible to learn several meanings for the same form when no phonological cues can be used to distinguish them? In the present work we provide a way out of this conundrum by showing that homophony is distributed non-arbitrarily in the lexicon, in a way that is compatible with children's available skillset during lexical acquisition. By allowing homophones in the language, but imposing that they are distributed in the lexicon so as to insure their distinctiveness, languages reach a trade-off between many different competing functional pressures.

To conclude, research in diachronic linguistics has suggested that homophones are dispreferred in languages and that this is reflected in preschoolers' failure to learn homophones in previous studies. Yet, this fails to explain the presence of homophony in languages, as well as how children eventually manage to learn these homophonous words. Our results bring elements of answers to these questions. We showed that children have no problem learning a pair of homophones, when each member of the pair appears in a different syntactic or semantic context, that indicate to children that a meaning distinction is necessary. This behavior is in line with the structure of the lexicon: Homophones are distributed in such a way that they spread out along the same dimensions that improve their learnability, beyond what would be expected by chance. We propose that learning (and possibly other functional constraints) exercises a finer-grained influence on the distribution of homophones in the lexicon, by selecting homophones whose meanings can be easily disambiguated by the context in which they occur, while pruning out homophones which are not distinguishable through their context, which makes them both hard to learn and prone to triggering misunderstandings.

Acknowledgments

This work was funded by a Post-Doctoral fellowship from the Fyssen Fondation and a ESRC Future Research Leaders award (ES/N017404/1) to ID. It was supported by grants from the French Agence Nationale de la Recherche (grants nos. ANR-13-APPR-0012, ANR-10-LABX-0087 IEC and ANR-10-IDEX-0001-02 PSL) and from the Fondation de France (grant no. 2012-00033704).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cogpsych.2018.04.001.

References

Altmann, G., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247–264.
André, M. (1955). *Economie des changements phonétiques.* Berne: Francke.
Arias-Trejo, N., & Plunkett, K. (2013). What's in a link: Associative and taxonomic priming effects in the infant lexicon. *Cognition, 128*(2), 214–227.
Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database (release 2)*. Distributed by the Linguistic Data Consortium, University of Pennsylvania.
Barkal, M. (1978). Phonological opacity v. semantic transparency: Two cases from Israeli Hebrew. *Lingua, 44*(4), 363–378.
Bates, E., & MacWhinney, B. (1982). Functionalist approaches to grammar.
Bernal, S., Dehaene-Lambertz, G., Millotte, S., & Christophe, A. (2010). Two-year-olds compute syntactic structure on-line. *Developmental Science, 13*(1), 69–76.
Bloomfield, L. (1933). *Language history: From Language* (1933 ed.). Holt, Rinehart and Winston.
Brusini, P., Dehaene-Lambertz, G., Dutat, M., Goffinet, F., & Christophe, A. (2016). ERP evidence for on-line syntactic computations in 2-year-olds. *Developmental Cognitive Neuroscience, 19*, 164–173.
Brusini, P., Dehaene-Lambertz, G., Van Heugten, M., de Carvalho, A., Goffinet, F., Fiévet, A.-C., & Christophe, A. (2017). Ambiguous function words do not prevent 18-month-olds from building accurate syntactic category expectations: An ERP study. *Neuropsychologia*.
Carey, S. (1978). The child as word learner (pp. 264–293).
Casenhiser, D. M. (2005). Children's resistance to homonymy: An experimental study of pseudohomonyms. *Journal of Child Language, 32*(2), 319–343.
Cassidy, K. W., Kelly, M. H., & Sharoni, L. A. J. (1999). Inferring gender from name phonology. *Journal of Experimental Psychology: General, 128*(3), 362.
Cauvet, E., Limissuri, R., Millotte, S., Skoruppa, K., Cabrol, D., & Christophe, A. (2014). Function words constrain on-line recognition of verbs and nouns in French 18-month-olds. *Language Learning and Development, 10*(1), 1–18.
Chomsky, N. (2002). An interview on minimalism. N. Chomsky, On Nature and Language, 92–161.
Chomsky, N. (1965). Aspects of the theory of syntax (Vol. 11). MIT Press.
Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences, 31*(05).
Coady, J. A., & Aslin, R. N. (2003). Phonological neighbourhoods in the developing lexicon. *Journal of Child Language, 30*(2), 441–469.
Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition, 106*(2), 633–664.
Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*(4), 148–153.
Culbertson, J., Gagliardi, A., & Smith, K. (2017). Competition between phonological and semantic cues in noun class learning. *Journal of Memory and Language, 92*, 343–358.
Dahan, D., Swingley, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic gender and spoken-word recognition in French. *Journal of Memory and Language, 42*(4), 465–480.
Dautriche, I., Chemla, E., & Christophe, A. (2016). Word learning: homophony and the distribution of learning exemplars. *Language Learning and Development, 12*(3), 231–251.
Dautriche, I., Mahowald, K., Gibson, E., Christophe, A., & Piantadosi, S. T. (2017). Words cluster phonetically beyond phonotactic regularities. *Cognition, 163*, 128–145.
Dautriche, I., Swingley, D., & Christophe, A. (2015). Learning novel phonological neighbors: Syntactic category matters. *Cognition, 143*, 77–86.
Dautriche, I. (2015). Weaving an ambiguous lexicon. Unpublished PhD thesis. Paris: Université Paris Descartes.
de Carvalho, A., Dautriche, I., & Christophe, A. (2016). Preschoolers use phrasal prosody online to constrain syntactic analysis. *Developmental Science, 19*, 235–250.
de Carvalho, A., Dautriche, I., Lin, I., & Christophe, A. (2017). Phrasal prosody constrains syntactic analysis in toddlers. *Cognition, 163*, 67–79.
Demuth, K., & Tremblay, A. (2008). Prosodically-conditioned variability in children's production of French determiners. *Journal of Child Language, 35*(1), 99–127.
Doherty, M. J. (2004). Children's difficulty in learning homonyms. *Journal of Child Language, 31*(1), 203–214.
Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41*(4), 469–495.
Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language. *Developmental Psycholinguistics: On-line Methods in Children's Language Processing, 44*, 97.
Gagliardi, A., & Lidz, J. (2014). Statistical insensitivity in the acquisition of Tsez noun classes. *Language, 90*(1), 58–89.
Hahn, N., Snedeker, J., & Rabagliati, H. (2015). Rapid linguistic ambiguity resolution in young children with autism spectrum disorder: Eye tracking evidence for the limits of weak central coherence. *Autism Research, 8*(6), 717–726.
Johnson, E. K. (2005). Grammatical gender and early word recognition in Dutch. In Proceedings of the 29th annual Boston University conference on language development (Vol. 1, pp. 320–330).
Kaplan, A. (2011). How much homophony is normal? *Journal of Linguistics, 47*(3), 631–671.
Karmiloff-Smith, A. (1981). A functional approach to child language: A study of determiners and reference (Vol. 24). Cambridge University Press.
Ke, J. (2006). A cross-linguistic quantitative study of homophony. *Journal of Quantitative Linguistics, 13*(01), 129–159.
Kedar, Y., Casasola, M., & Lust, B. (2006). Getting there faster: 18- and 24-month-old infants' use of function words to determine reference. *Child Development, 77*(2), 325–338.
Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review, 99*(2), 349.
Kern, S. (2007). Lexicon development in French-speaking infants. *First Language, 27*(3), 227–250.
Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*(2), 211.
Lee, C., & Federmeier, K. D. (2009). Wave-ering: An ERP study of syntactic and semantic context effects on ambiguity resolution for noun/verb homographs. *Journal of Memory and Language, 61*(4), 538–555.
Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science, 18*(3), 193–198.
Li, C. N., & Thompson, S. A. (1989). Mandarin Chinese: A functional reference grammar. Univ. of California Press.
Neighborhoods of Words in the Mental Lexicon. Research on Speech Perception. Technical Report No. 6.
Mahowald, K., Dautriche, I., Gibson, E., & Piantadosi, S. T. (in review). Word forms are structured for efficient use.
Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177–190.
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.
Millotte, S., René, A., Wales, R., & Christophe, A. (2008). Phonological phrase boundaries constrain the online syntactic analysis of spoken sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(4), 874.

New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers, 36*(3), 516–524.

Newman, R., Samuelson, L., & Gupta, P. (2008). Learning Novel Neighbors: Distributed mappings help children and connectionist models. In 30th annual meeting of the cognitive science society.

Nieuwland, M. S., & Van Berkum, J. J. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience, 18*(7), 1098–1111.

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., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition, 122*(3), 280–291.

Quine, W. V. O. (1960). *Word and Object.* MIT Press.

Řehůřek, R., Sojka, P., & others. (2010). Software framework for topic modelling with large corpora.

Setoh, P., Wu, D., Baillargeon, R., & Gelman, R. (2013). Young infants have biological expectations about animals. *Proceedings of the National Academy of Sciences, 110*(40), 15937–15942.

Shi, R., & Melancon, A. (2010). Syntactic Categorization in French-Learning Infants. *Infancy, 15*(5), 517–533.

Slobin, D. I. (1985). Crosslinguistic evidence for the language-making capacity. In D. I. Slobin (Ed.), The crosslinguistic study of language acquisition (Vol. 2). Hillsdale, NJ: Erlbaum.

Slobin, D. I. (1978). Cognitive prerequisites for the development of grammar. CA Ferguson & D, 1.

Spinelli, E., & Alario, F. X. (2002). Gender context effects on homophone words. *Language and Cognitive Processes, 17*(5), 457–469.

Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research, 44*(6), 1321–1337.

Storkel, H. L., & Hoover, J. R. (2011). The influence of part-word phonotactic probability/neighborhood density on word learning by preschool children varying in expressive vocabulary. *Journal of Child Language, 38*(03), 628–643.

Storkel, H. L., & Lee, S.-Y. (2011). The independent effects of phonotactic probability and neighbourhood density on lexical acquisition by preschool children. *Language and Cognitive Processes, 26*(2), 191–211.

Storkel, H. L., & Maekawa, J. (2005). A comparison of homonym and novel word learning: The role of phonotactic probability and word frequency. *Journal of Child Language, 32*(4), 827.

Storkel, H. L., Maekawa, J., & Aschenbrenner, A. J. (2013). The effect of homonymy on learning correctly articulated versus misarticulated words. *Journal of Speech, Language, and Hearing Research, 56*(2), 694–707.

Swingley, D., & Aslin, R. N. (2007). Lexical competition in young children's word learning. *Cognitive Psychology, 54*(2), 99.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*(5217), 1632–1634.

Van Heugten, M., & Christophe, A. (2015). Infants' acquisition of grammatical gender dependencies. *Infancy, 20*(6), 675–683.

Van Heugten, M., & Shi, R. (2009). French-learning toddlers use gender information on determiners during word recognition. *Developmental Science, 12*(3), 419–425.

Veneziano, E., & Parisse, C. (2010). The acquisition of early verbs in French: Assessing the role of conversation and of child-directed speech. *First Language, 30*(3–4), 287–311.

Von Holzen, K., & Mani, N. (2012). Language nonselective lexical access in bilingual toddlers. *Journal of Experimental Child Psychology, 113*(4), 569–586.

Wasow, T., Perfors, A., & Beaver, D. (2005). *Morphology and the web of grammar: Essays in memory of Steven G. Lapointe.* CSLI Publications.

Wedel, A., Jackson, S., & Kaplan, A. (2013). Functional load and the lexicon: evidence that syntactic category and frequency relationships in minimal lemma pairs predict the loss of phoneme contrasts in language change. *Language and Speech, 56*(3), 395–417.

Wedel, A., Kaplan, A., & Jackson, S. (2013). High functional load inhibits phonological contrast loss: A corpus study. *Cognition, 128*(2), 179–186.

Zangl, R., & Fernald, A. (2007). Increasing flexibility in children's online processing of grammatical and nonce determiners in fluent speech. *Language Learning and Development, 3*(3), 199–231.