A perceptual advantage for onomatopoeia in early word learning: Evidence from eye-tracking

Abstract

A perceptual advantage for iconic forms in infant language learning has been widely reported in the literature (e.g. Asano et al., 2015), termed the 'sound symbolism bootstrapping hypothesis' by Imai and Kita (2014). However, empirical research in this area is limited mainly to sound symbolic forms, which are very common in languages such as Japanese, but less so in Indo-European languages such as English. In this study, we extend this body of research to onomatopoeia – words which are thought to be present across most of the world's languages (Hinton et al., 1994), and which are known to be dominant in infants' early lexicons (e.g. Tardif et al., 1998). In a picture-mapping task, 10- and 11-month-old infants showed a processing advantage for onomatopoeia (e.g. *woof woof*) over their conventional counterparts (e.g. *doggie*). However, further analysis suggests that the input may play a key role in infants' experience and processing of these forms.

Keywords: iconicity, language acquisition, eye-tracking, onomatopoeia, early language processing

A large body of recent evidence has consistently shown an advantage for iconicity – that is, forms that are symbolically linked to their meanings through seemingly non-arbitrary verbal or gestural cues – in language learning (Asano et al., 2015; Imai et al., 2008; Lockwood et al., 2016). This is consistent across a range of populations, using a wide spectrum of experimental designs. However, in infant research in particular, the stimuli used in these various experiments are not typical of the real-world language-learning experience. Most often, participants are tested on their ability to pair non-words with novel objects (e.g. Asano et al., 2015, Ozturk et al., 2013), or their success in learning novel non-words which are perhaps atypical of the native language in question (Imai et al., 2008). Therefore, while results suggest an inherent sensitivity to iconic

sound-meaning correspondences across populations, it is difficult to extend the validity of these findings to our understanding of language learning in reality. In this study, we address this gap in the literature by testing whether young infants show a learning preference towards iconic words that are present in their native language. Eye-tracking is used to analyse infants' processing of onomatopoeia – a form of iconicity that is present in most, if not all, languages of the world (Hinton et al., 1994), and which is considered in some accounts to be an example of 'true' sound symbolism (Lyons, 1968; Sapir, 1970).

Over the last decade, an increasing body of research has considered whether non-arbitrariness in language might facilitate learning. In contradiction with the Saussurean notion of 'the arbitrariness of the sign' (1962), it has been posited that language may be more easily acquired across forms which are somehow iconically connected to their meaning (Imai & Kita, 2014). This advantage has been demonstrated in signed languages (Vinson et al., 2008), where signs rated as more iconic tend to be acquired earlier than less iconic signs. Furthermore, studies of both adults and toddlers have shown a learning advantage for sound symbolism in spoken language: non-Japanese speakers learn sound symbolic Japanese adjectives (Lockwood et al., 2016) and novel verbs (Imai et al, 2008) more successfully than non-sound symbolic words. The established evidence appears to support Imai and Kita's (2014) sound symbolism bootstrapping hypothesis, as learners do indeed appear to draw upon iconicity to facilitate language learning. However, the extent to which different languages show iconic form-meaning correspondences is still unclear, and thus the broader relevance of these findings to language acquisition in general should not be taken for granted.

Certainly, non-arbitrariness is identifiable across many languages. Onomatopoeia (e.g. *woof woof*, *bang*) probably constitute the most obvious and common form of iconicity in language as a whole, but ideaphones (*glisten*, *jingle*) and mimetics (typically found in Japanese or Korean, including *gorogoro* 'movement of a heavy object' and *pika* 'a flash of light', Kita, 1997) also

contribute to iconicity in the linguistic system, albeit in varying proportions across languages. Typically, the research in this area has focussed on sound symbolism, or what Ohala (1984) termed the 'frequency code'; that is, a symbolic relationship between the formant values and vocal tract size in the production of a specific segment, and its corresponding meaning. Vowels and consonants with a smaller vocal tract size, and thus higher f0, such as /i/ and /k/, refer to small, sharp or rapid referents, while those produced with a larger space in the vocal tract and lower f0 such as /u/ and /b/ relate to large, slow or heavy referents (Hinton et al., 1994). Mimetics are also derived from these correspondences (Ivanova, 2006); Kita (1997, p.380) claims that "in [the] realm of mimetic forms, phonemes seem to have meanings of their own".

A multitude of experiments have used these correspondences, known as the 'bouba/kiki' effect (Köhler, 1970), to test effects of sound symbolism. In this paradigm, round and spiky shapes are presented with either congruent or incongruent 'round' or 'spiky' words, corresponding to labels such as *bouba* and *kiki*, respectively. Evidence shows that sensitivity to these correspondences is consistent across speakers of different languages and age-ranges (Davis, 1961), as well as with both novel and familiar objects (D'Onofrio, 2014). It has been suggested that this sensitivity to sound-symbol correspondences might be an innate aspect of primate cognition (Ozturk et al., 2013), since infants as young as four months match congruent sound-shape correspondences. The suggestion that non-arbitrary form-meaning correspondences may be innately specified is supported by findings from Bohn and colleagues (2016), who showed that even chimpanzees draw upon iconic gestures when learning associations between form and meaning. However, the chimpanzees were much slower to learn than four-year-old children, who were consistent in drawing upon iconicity in their comprehension of novel gestures.

While these studies may demonstrate important findings regarding the nature of our sensitivity to sound symbolic congruence between form and meaning, it is difficult to extend their conclusions to real-world situations of learning. To what extent do such form-meaning

correspondences occur across languages in general, and are these correspondences ubiquitous in supporting learning regardless of the language being acquired? Recent literature has helped to broaden our understanding of these issues; it seems that even languages which are generally considered to contain only few examples of iconicity may be more systematically sound symbolic than they initially appear. Monaghan and colleagues (2014) analysed a corpus of English data to show that the most systematic sound-meaning correspondences were typical of words acquired earlier in development. Their study only considered the lexicon from age three, but Perry, Perlman and Lupyan's (2015) analysis of systematicity and iconicity across around 600 early-acquired words in English and Spanish (as reported on the MacArthur Bates Communicative Development Inventories (CDI), Fenson et al., 1994) confirmed this to be the case. From the very first stages of infant production, the forms produced earliest tended to have the most systematic form-meaning correspondences. Importantly, their results also showed that the earliest-acquired words were those judged as being most iconic. Unsurprisingly, onomatopoeia and interjections were consistently rated as being highest in iconicity, and these forms are often acquired in infants' very earliest productive vocabularies. However, the effect of iconicity on age of acquisition remained even when onomatopoeia and interjections were removed from the data. These results present strong evidence towards an advantage for iconic forms in early word learning across both English and Spanish, and may explain the high proportion of onomatopoeia in infants' earliest words (e.g. Tardif et al., 2008). However, such considerations must be taken with caution. While iconic forms may be more commonly acquired in the very earliest stages of word production, this does not automatically point to a language-general processing advantage for iconicity. A closer look at early infant data (as reported on the CLEX database, Jørgensen et al., 2010) demonstrates that this cannot be the only factor motivating the early acquisition of a word form. The average iconicity rating (Perry et al., 2015) of the ten words produced most

frequently by American 12-month-olds is 1.94, and ranges from -0.07 (*dog*) to 3.44 (*yum*).[1] If iconicity were the sole driving factor behind acquisition, we would perhaps expect to see a higher average rating overall. Furthermore, when we attempt to consider the systematicity of these first ten words, only two– *dog* and *ball* – are included in the data reported in both Monaghan and colleagues' (2014) and Perry and colleagues' (2015) analyses. These two words are amongst the three least iconic of the ten words overall, and *dog* is both the least iconic and the least systematic of the two. In these two cases, therefore, the motivation behind their early acquisition cannot be driven by iconicity. While iconicity may be an important factor in early language learning overall, and may play an important role in the early establishment of form-meaning correspondences (Imai & Kita, 2014), it does not necessarily lead infants to the early production of these forms: that is, we cannot draw conclusions about the role of iconicity in early language learning from production data alone. In order to better understand the role that iconicity plays in the development of early form-meaning mappings, it is essential to consider the early vocabulary (i.e. those words which tend to be produced first – both iconic and non-iconic) alongside infants' on-line processing of these forms. In this way we can more clearly establish the role of iconicity and its potential bootstrapping effects in early language learning, independent of the limitations and preferences presented by memory and articulatory capacities in vocal production.

Experimental evidence in this domain is biased towards the iconic correspondences typical of Japanese, and as such, the relevance of iconicity to early language processing cannot yet be extended to all languages. Japanese infants have been found to demonstrate sensitivity to sound symbolism in non-words from a young age (Asano et al., 2015), and similar experiments have generated parallel findings across American four-month-olds (Ozturk et al., 2013), Spanish four-month-olds (Peña et al., 2011), and American toddlers (Maurer et al., 2006), among others. While this body of research suggests an early processing advantage for the sound-meaning

[1] Adults were asked to rate each word on a scale of -5 ('sounds like the opposite of what it means') to 5 (sounds like what it means; Perry et al, 2015).

correspondences typically found in mimetic forms, we must question the extent to which this is relevant to early word learning across languages. One might suggest that the sound symbolism bootstrapping hypothesis (Imai & Kita, 2014) is relevant to the acquisition of mimetic languages such as Japanese and Korean, but perhaps not extendable as a language-general phenomenon.

In order to test the relevance of the sound symbolism bootstrapping hypothesis to word learning in general, we must extend our analysis to types of iconicity that can be found across languages. One aspect of iconicity that is consistently referred to in relation to infant language is onomatopoeia. However, despite the broadening literature in iconicity research, onomatopoeia have been largely overlooked in empirical studies. Regularly reported as being prominent in infants' earliest words (Menn & Vihman, 2011; Tardif et al., 2008), onomatopoeia are thought to pose a problem to researchers trying to understand the earliest stages of language production, and as a consequence they are often removed from analyses of production data (e.g. Behrens, 2006; Fikkert & Levelt, 2008). In her analysis of syllabification in Finnish infants' early words, Kunnari (2002) chooses to analyse onomatopoeia separately from the rest of the infants' lexicons, as she observes that the high proportion of these forms (24% of the output at the 15-word point) distort her overall results. It seems unclear as to whether onomatopoeia should be considered as part of the developing lexicon, or whether they are "nonce words" to be disregarded alongside "hesitation markers…and noninterpretable words" (Behrens, 2006, p.15). However, at the same time, onomatopoeia are claimed to be "one of the most obvious examples" (Nygaard et al., 2009, p.181) of non-arbitrary sound-symbol correspondences, and are discussed as part of Imai and Kita's (2014) sound symbolism bootstrapping hypothesis.

In this study, we test the effect of iconicity in early language perception using onomatopoeia as our stimuli. This will make it possible to observe whether the sound symbolism bootstrapping hypothesis (Imai & Kita, 2014) extends to onomatopoeia. Positive results would lead us to

reconsider the role of onomatopoeia in early language learning, broadening our understanding of how iconicity may relate to real-world acquisition across languages.

We address the question of iconic bootstrapping in the acquisition of onomatopoeic words (OWs) using eye-tracking. In a picture-mapping task, 10-month-old infants were tested on their ability to match OWs and their corresponding conventional word (CW) forms to their referent. Fixation time and response latency were measured to determine whether there is any perceptual advantage for onomatopoeia in early language learning. If the iconic advantage is present in onomatopoeia as it is found to be in studies of sound symbolism, infants should be better able to match words to images in the OW condition than the CW condition. This would reflect findings from Asano and colleagues (2015), among others, who have shown that infants are sensitive to sound symbolic form-meaning correspondences from a young age.

Methodology

Participants

Parents were recruited through an advertisement in a local magazine, through social media and by word of mouth. Forty 10- and 11-month-old infants took part in the experiment overall, of which 13 were excluded from the analysis due to equipment failure (n= 5), fussiness (n= 3) and calibration problems (n= 4), leaving a total of 27 infants for analysis (14 females, mean age 328 days). All infants were acquiring Northern British English typical of York, UK, where the experiment was carried out. One infant was receiving some exposure to Mandarin alongside English. No developmental difficulties were reported, and all but two infants had full-term gestational periods. The two infants in question were dizygotic twins; the overall results did not differ when these infants' data were removed from the analysis, and so we included them in the analyses reported here.

Stimuli and materials

Audio stimuli.

Six OWs from the 'sound effects and animal sounds' section of the Oxford Communicative Development Inventory (CDI, Hamilton et al., 2000) were selected for use in the test trials, each with a CW counterpart from elsewhere on the CDI. These words were chosen on the basis that they would be familiar to 10 and 11-month-old infants. Two further OW-CW pairs from the CDI were selected as filler trials. Stimuli are detailed in Table 1; all OWs matched our definition of onomatopoeia as conventionalized vocal imitations of sounds from the environment. Audio stimuli were recorded by a female speaker of Northern British English. The speaker – a linguist from the department at York – was asked to produce each word in the carrier phrase 'Where's the [target]?'. It was specified that target words should be produced with matching pitch contours and in typical infant-directed speech style, but with no specific sound effects. Word duration ranged from .524ms to .824ms across OW and CW stimuli.

Table 1: OW and CW stimuli used in the experiment.

OW	CW
Baa	Sheep
Meow	Kitty
Moo	Cow
Vroom	Car
Quack quack	Ducky
Woof woof	Doggie
Cock-a-doodle-doo	*Cockrell*
Choo choo	*Train*

Filler stimuli are marked in italics.

Multiple tokens of each target were recorded, and the final stimuli were selected based on the closest match in pitch and duration for each OW-CW pair to avoid a bias based on prosodic

salience. We ran a linear regression model in R (R core team, 2016) to verify that none of the stimuli stood out as more salient than the others; mean pitch, pitch range and duration were included as dependent variables, with type (OW vs. CW) as a fixed effect. No significant differences were found for any of the three measures (mean pitch: t=-.002, p=.998; pitch range: t=2.05, p=.068; duration: t=.922, p=.378). Each test pair was also matched for number of syllables; this was not possible in the filler pairs but was not considered to be an issue. For three of the stimuli – CAT, DOG, and DUCK – diminutives were selected for the CW in each pair, as this was considered to be a common production of these target words. This was matched to the syllabification of the corresponding OWs, with the inclusion of reduplication in *quack quack* and *woof woof*.

As this experiment is designed to test infants' perception of iconic versus non-iconic stimuli, we compared iconicity ratings across the stimuli to make sure that OWs were indeed interpreted as being more iconic[2]. Iconicity ratings of spoken words used in Perry and colleagues' (2015, supporting information S1) analysis were compared across stimuli using a linear regression model, with iconicity ratings from -5 ("words that sound like the opposite of what they mean") to 5 ("words that sound like what they mean", Perry et al., 2015) as the dependent variable and Type (OW vs. CW) and stimuli as fixed effects. A highly significant effect was found for type (p<.001, t=7.42); on average OWs were rated as being 3.13 points higher in iconicity, according to Perry and colleagues' rating scale. This confirmed that the comparison of OW and CW stimuli in this experiment would be contrasting perception of more- versus less-iconic stimuli. However, it is important to note that the CW stimuli – while being rated as less iconic than their OW counterparts – were not necessarily considered to be 'non-iconic'. The CW stimuli generated a mean iconicity rating of .57 (SD=.75, OWs: M=3.7, SD=.45), pointing to some interpretation of iconicity across all stimuli.

[2] With thanks to an anonymous reviewer for this suggestion.

Visual stimuli.

Two photographic images were selected for each target, with one image for each of the filler pairs. The six animals all stood facing towards the right-hand side of the screen with their head turned towards the infant, while the car and train images were presented with the front-end of the vehicle towards the right-hand side of the screen. Two images of approximately 400x400 pixels in size were presented in colour, side-by-side, on a 1280x1024 pixel grey background. These were vertically centralised, and set at a distance of 120 pixels from the edge of the screen on either side.

Apparatus

The experiment was controlled using PsychoPy Experiment Builder (Peirce, 2007) and was run through Tobii Studio, presented on a 17" Tobii Studio T60 eye-tracking monitor sampling at 60Hz. The experiment was set up in a darkened sound-attenuated booth with speakers installed in the walls of the booth at each side of the monitor, and stimuli were played at a volume of approximately 70dB. A video camera positioned above the eye-tracker allowed the experimenter to view the procedure from the adjoining room, providing information regarding the infants' comfort as well as orientation towards the screen.

Procedure

An image from the children's TV programme *Teletubbies* was displayed on the screen prior to the infant entering the experimental booth. This distracted the infant while the experimenter set up the procedure and drew their attention towards the screen ready for the start of the experiment. Caregivers wore foam earplugs and headphones playing multi-talker babble as well as visibility-blocking glasses, to ensure no influence on infants' responses during the experiment. Infants sat on their laps in a chair placed 60cm from the screen.

A five-point infant calibration was taken, after which point the experiment began. This lasted approximately 3.5 minutes and consisted of 24 pseudo-randomised test trials in four phases. A filler trial was presented at the start of each phase, followed by six test trials. Each infant heard 14 OWs and 14 CWs, presented in a random order. The display was counterbalanced so that the target image appeared equally on both the left- and right-hand sides of the screen.

Infants were presented with two images in a 3700ms familiarisation phase before the audio stimuli ('Where's the [target]?') was played. The images remained on screen until 3500ms after the onset of the audio stimuli. After the experiment infants were rewarded with a York Babylab t-shirt and parents were asked to complete the Oxford CDI questionnaire (Hamilton et al., 2000).

Results

Data were analysed from a window of 350 to approximately 2500ms after onset of the audio stimuli. This analysis period was selected based on Swingley and colleagues' approach (Swingley, Pinto & Fernald, 1999; Swingley & Aslin, 2000), who typically adjust the assumed latency period between target onset and response according to the age of the infant. Twelve-month olds have been found to show mean saccade latencies of around 290ms (Canfield et al, 1997), and so this was extended here to allow a slightly longer response time for 10-11 month old infants. The offset of the analysis window was also adjusted from that of Swingley and Aslin's (2000) study in order to allow for slower and longer responses from the younger infants.

Proportion of looking towards the target image was calculated for each trial as a percentage of the total fixation time for both target and distractor, and a mean looking time was calculated for each infant for both OW and CW stimuli across the six targets. Five infants were excluded from the analysis due to a lack of matching data across OW and CW conditions for any stimuli (i.e. responded to CW *doggie* but not to OW *woof woof*). This left 22 infants for analysis. In order to

assess overall recognition of the stimuli, mean difference scores (MDS) were calculated across the results (Bergelson & Swingley, 2012). Differences in fixation proportions were used as an indication of whether the infants recognised the six stimuli and their corresponding OW and CW labels: this was calculated as the difference in fixation proportion to an image as target in relation to the fixation proportion to the same image as a distractor (see Bergelson & Swingley, 2012). Eighteen of the 22 infants showed a positive MDS across the results (M=.045, SD=.15), and binomial tests confirmed this to be significantly above chance (p=.004). This rules out the possibility of a preference to any of the six individual targets.

Of the data from the 22 infants, 63% of trials – 355 overall – provided useable eye-tracking data. We considered both proportion of fixations to the target image and response latency as dependent measures in the analysis. Shapiro-Wilk tests confirmed normality across both OWs and CWs for fixation proportions (OW: p=.827, CW: p=.346) and response latencies (OW: p=.187, CW: p=.596), so parametric tests were used throughout the analysis. Effects of age, sex and language ability (number of words the infant was reported to produce, according to CDI questionnaires) were tested using a linear regression model in R. Infants' production capacity ranged from zero to five words (M=1.4, SD=1.9); the majority of infants (n=10) produced no words, while three infants were reported to produce five words. Mean fixation proportion was included as the dependent variable, with age, sex and language ability as fixed effects. None of these factors had any effect the infants' overall mean fixation proportion (age: t=-.86, p=.39, sex: t=1.15, p=.25, language ability: t=-.36, p=.72) and so were not considered in further analyses.

Next we analysed fixation proportions to compare infants' looking behaviour across OW and CW conditions. A linear mixed-effects model was generated with fixation to the target image as the dependent variable and the fixed effects of condition (OW vs. CW) and stimuli type (CAR, CAT, COW, DOG, DUCK or SHEEP). Subject was included as a random effect, with by-subject random slopes for the effect of condition. Infants fixated 8% longer in the OW condition, and

likelihood ratio tests showed this difference to be significant ($\chi2$ (1)=4.31, p=.038, t=2.09). A significant effect was also found for stimuli type ($\chi2$ (1)=11.87, p=.036); infants fixated longest when the target was CAT, COW or DOG (all ts<2), and marginally also for DUCK targets (t=1.9). This is shown in Figure 1b. However, as Figure 1a suggests, infants were not looking to the OW target significantly above chance overall (p=.2).

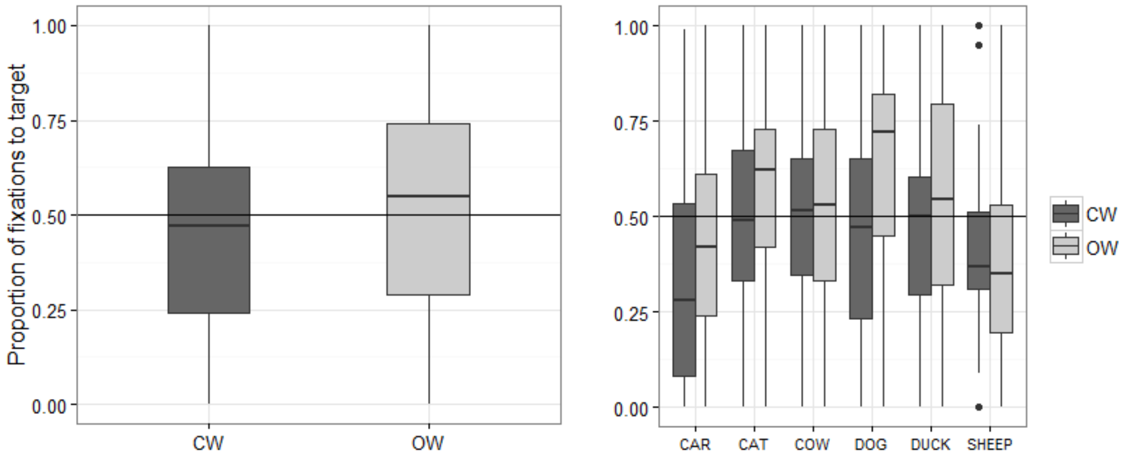

Figure 1a-b: Total fixation proportion to target across OW and CW stimuli.

On average, overall looking was below chance in the CW condition (M=.45), and CAR and SHEEP showed below-chance looking in both OW and CW conditions (CW CAR: M=.36, CW SHEEP: M=.42; OW CAR: M=.42, OW SHEEP: M=.39).

We then analysed response latencies across OW and CW forms. Trials in which the infant was fixating on the distractor image at the onset of the test phase were extracted and response latency from the distractor to the target image was measured and compared across the two conditions. As explained for the overall analysis above, infants were removed from the analysis if their data did not include any response latency measures for both OW and CW conditions across stimuli. Only one infant did not provide enough data across conditions, while a second infant demonstrated anomalously long response times across trials (see below), leaving 20 infants with useable response latency data. This left us with 107 trials for the response latency analysis,

accounting for 32% of trials overall. Outliers were then removed according to approaches reported by Fernald and colleagues (1998, 2008), adjusted to meet the looking behaviour of 10-month-old infants. Responses faster than 150ms or slower than 2000ms were excluded from the dataset, as these were considered unlikely to reflect a response to the target word. This excluded data from 27 trials in the final analysis leaving 80 trials overall.

A linear mixed-effects model was generated with time to first fixation on the target image as the dependent variable and fixed effects of condition and stimuli type; subject was included as a random effect, with by-subject random slopes for the effect of condition. Likelihood ratio tests showed no effect for stimulus type on infants' response latencies ($\chi 2$ (1)=.536, p=.46, t=-.76). On average, responses were 85ms faster in the OW condition, but variability was high across infants (M=994ms, SD=516ms) and across stimuli (Figure 2a-b). The same model showed no effect for stimulus type on response latency ($\chi 2$ (1)=8.23, p=.14).

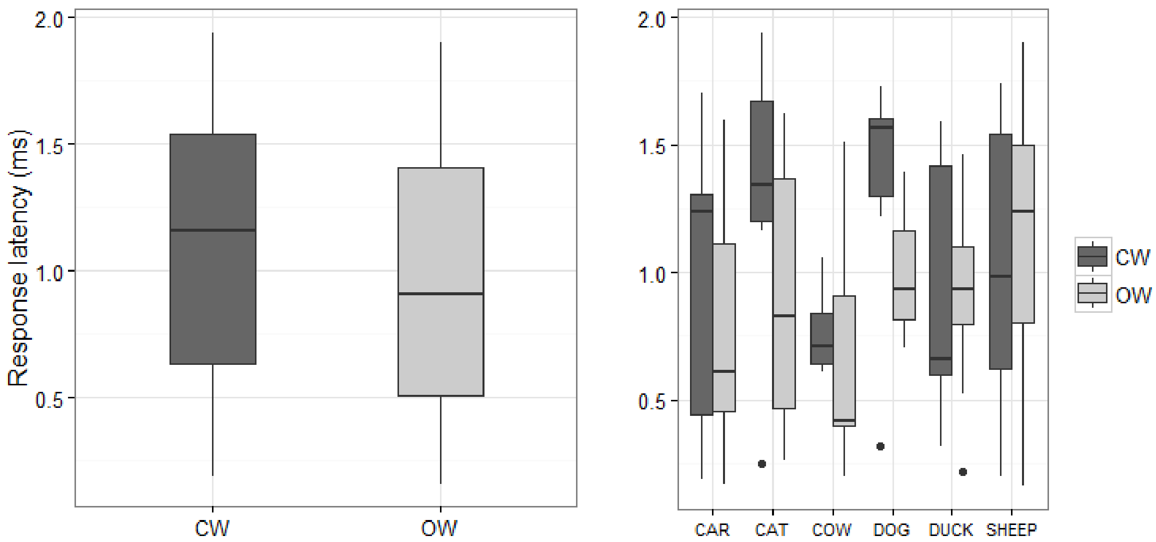

Figure 2a-b: Mean response latencies across OW and CW stimuli.

So far our results show a learning bias towards onomatopoeia in early language perception, but the fact that OW forms do not generate above chance overall looking calls this into question. We

investigated this further using the CDI reports to determine whether infants' understanding of the individual stimuli was affecting their looking behaviour.

Of the six stimuli, DOG was reported to be the most understood while CAR was the least understood: 62% of infants were reported to understand either OW or CW DOG or both, compared to only 19% of infants for CAR. This contrasts slightly with language acquisition norms of American infants from the CLEX database (Jørgensen et al., 2010), where DOG is the most understood word at 10 months (OW: 35.8%, CW: 51.1%), while CW *sheep* (4.4%) and OW *moo* (13.2%) are reported as the least understood words. Across participants, only four of the stimuli were reported as being spoken, across three infants' data; these were all CWs – no OWs were reported to be spoken by any of the infants. Again this contrasts with norms reported in the CLEX database, where a mean of 8.2% of infants produce our OW stimuli at 10 months, compared with 2.7% for the CW stimuli. However, both the CLEX norms and our CDI reports show no difference in understanding of the OW and CW forms: OWs accounted for 48% of the 'understood' words on the CDI reports, while mean understanding of the CW and OW forms was almost identical in the CLEX data from 10-month-olds (OW: 25.9%, CW: 25.2%). Overall, these comparisons suggest no overt bias towards OW acquisition. We tested both of these measures on the infants' processing time by including them as factors in our model. First we tested acquisition norms (word production at 10 months) of each of the stimuli, with proportion of fixations to target as the dependent variable, condition and stimuli type as fixed effects, and subject as a random effect, with random slopes for the effect of stimuli type. Infants did not fixate longer to the stimuli which are most often produced by 10-month-olds according to the CLEX database ($\chi2(1)=1.61$, $p=.21$, $t=-1.27$). Reported understanding (understands vs. doesn't understand) according to the infants' own CDIs was then tested in the same model, and again this had no effect on the infants' responses ($\chi2(1)=.313$, $p=.58$, $t=-.56$). The same two models were then run with response latency as the dependent variable, and again no effect was found for

acquisition norms ($\chi2(1)=2.45$, $p=.12$, $t=1.58$), nor for reported understanding, though this was marginally significant ($\chi2(1)=3.56$, $p=.06$, $t=1.97$). The infants showed no bias towards those stimuli that are most commonly produced at 10 months by American infants, nor towards the stimuli that they were reported to understand on their CDI forms.

We then tested iconicity ratings as a factor in our model, to see if differences in perceived iconicity (as judged by adult listeners) affected infants' perception of the stimuli. We included iconicity rating as a fixed effect in our model, which did not have any effect on infants' fixation proportions ($\chi2(1)=2.1$, $p=.15$, $t=1.45$). When the same model was generated with response latency as the dependent variable, again no effect was found ($\chi2(1)=.05$, $p=.82$, $t=-.23$).

We then returned to the results shown in Figure 1b, where there appears to be a bias against the stimuli CAR and SHEEP. This corresponds to the parental reports; despite the statistics showing no effect for reported understanding on individual infants' responses overall, CAR and SHEEP (OW, CW, or both) were understood by only four and five of the infants, respectively. We tested this using a linear regression model with proportion of fixations as the dependent variable and stimuli type as a fixed effect, and a significant effect was found for stimuli type ($F(5, 241)=2.34$, $p=.04$). Infants fixated significantly longer on CAT, DOG and COW (all $ps<.03$) than on CAR and SHEEP, and marginally longer on DUCK ($p=.063$). This suggests that CAR and SHEEP – both reported as the least understood words across the stimuli – may have been distorting the results reported above.

To test this, we grouped the stimuli into two categories according to understanding – 'common' (CAT, COW, DOG and DUCK) and 'uncommon' (SHEEP and CAR) – in order to observe whether or not these stimuli were affecting the data at the group level. A two-way ANOVA with fixation proportion as the dependent variable and stimuli commonness (common vs. uncommon) and condition (OW vs. CW) as fixed effects showed a highly significant effect for stimuli commonness on looking proportion ($F(1, 243)=11.53$, $p<.000$) and a significant effect for

condition (F(1, 243)=.896, p=.041). No interaction was found (F(1, 243)=.896, p=.34). On average infants were attending to the 'common' stimuli 13% longer than the 'uncommon' stimuli. These results justify the exclusion of CAR and SHEEP from the data for re-analysis. No effect was found when the same model was tested with response latency as a dependent variable (stimuli commonness: F(1, 96)=.38, p=.54; condition: F(1, 96)=1.9, p=.85), stimuli commonness*condition: F(1, 96)=1.9, p=.17).

Proportion of fixations to target and response latency were then re-tested across these four stimuli. A significant effect was found for condition on fixation proportions ($\chi 2(1)=5.36$, p=.021, t=2.34, see Figure 3a); infants were looking on average 10% longer to the target image in the OW condition, and this was significantly above chance (p=.03). No effect was found for stimuli type across this dataset ($\chi 2(1)=.401$, p=.94, Figure 3b). The same model was then tested with response latency as a dependent variable and no effect was found for condition ($\chi 2(1)=.515$, p=.47, t=-.73) or stimuli type ($\chi 2(1)=7.16$, p=.067, ts all <-.26).

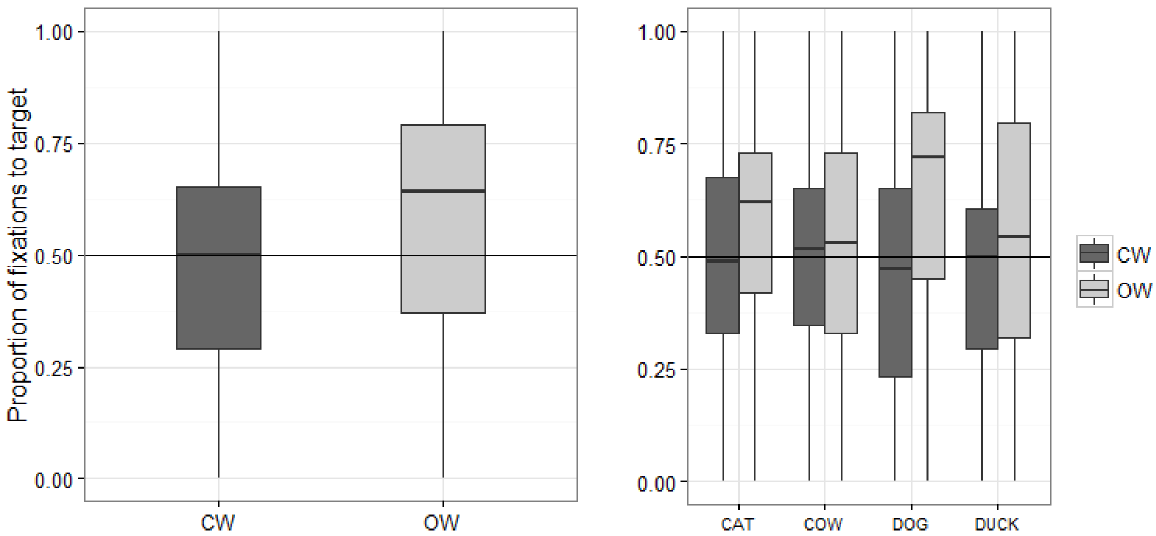

Figure 3a-b: Total fixation proportion to target across CAT, COW, DOG and DUCK stimuli only.

Discussion

This study set out to test whether infants' documented sensitivity to sound symbolism in early development can be extended to onomatopoeia. Previous studies have shown that infants can detect sound symbolic properties in non-words from as early as four months (Ozturk et al., 2013), and that by 11 months of age infants are quicker to map sound symbolic novel words to their referents than non-sound symbolic word-referent pairs (Asano et al., 2015). These results suggest a processing advantage for iconicity, but we questioned their relevance to language learning in reality. We hypothesised that if this sensitivity were functional in early language development – as proposed by the sound symbolism bootstrapping hypothesis (Imai & Kita, 2014) – then infants in the early stages of lexical acquisition should show more established lexical knowledge of iconic forms that are already present in the ambient language, namely onomatopoeia.

Infants were better able to match target images to their labels in the OW condition, but while this difference was significant, still the average looking across OW stimuli was below chance. However, results also demonstrated a significant effect for stimuli type. When we analysed this more closely, we found no effect for words which the infants were reported to understand on their individual CDI forms, but we did see a group tendency towards understanding of four of the stimuli in the fixation results. This corresponded to the grouped CDI data, which showed the same four words to be the most commonly understood across the infants. Statistical tests confirmed that the infants were indeed attending significantly longer to these forms, and a reanalysis of our results in line with these findings again showed infants to be fixating significantly longer in the OW condition. Furthermore, this time their responses were reliably above chance. No difference was found between infants' response latencies across the data – while infants tended to show shorter response latencies in the OW condition, this did not stand up to statistical testing, likely due to an insufficient amount of data across conditions.

Our results suggest that the effect found for sound symbolism in early language development may also be relevant to onomatopoeia – a feature of the infant (and adult) lexicon that is present in many of the world's languages, and that is particularly predominant in early infant speech. However, further testing is necessary if we are to draw any firm conclusions in this domain. In particular, an analysis of younger infants who have not yet begun to pair words with meanings would allow us to determine whether there is an automatic learning advantage for onomatopoeia, similar to that observed in experimental studies of sound symbolism (e.g. Ozturk et al., 2013). Alternatively, language experience gained in the first year of life may promote the preferential acquisition of iconic forms. A potential processing advantage for OWs even at 10 months of age may provide an explanation as to why infants acquire so many onomatopoeia in their earliest words, as has been reported across studies of early production (Menn & Vihman, 2011; Tardif et al., 2008). The iconic status of onomatopoeia may facilitate their early acquisition owing to their more transparent sound-meaning correspondences. Werner and Kaplan (1963) relate to the undifferentiated state of the internal and external forms – that is, a referent and its symbol – in early development, as young infants struggle to grasp abstract conceptual representations. For this reason, less abstract forms such as onomatopoeia are easier to represent in early acquisition, and thus tend to be among the first words to appear in the infant lexicon. This is supported in recent theoretical work by Dingemanse (2012), who highlights that onomatopoeia present particularly direct form-meaning mappings, as the sound of the human voice is used to represent sound from the environment. This sound-to-sound mapping may be pragmatically more salient in early language learning, as onomatopoeia function as forms that are both linguistic and performative in nature (Sasamoto & Jackson, 2016).

It is likely that numerous factors contribute to infants' early acquisition of specific word forms, as has previously been reported in the literature for onomatopoeia. A study by [blinded for review] (2014) observes the word-by-word acquisition of a German infant, Annalena (Elsen,

1991), highlighting how, of the 16 onomatopoeia in Annalena's early lexicon, all instances of OWs are acquired *before* the corresponding CW. This points to a learning advantage for OWs, which may be manifested in early perception, as well as production. [blinded for review] (2014) shows that OWs present a more suitable match to the common production patterns found throughout Annalena's early words (Vihman, 2016): the results suggest that OWs are acquired preferentially owing to their suitability for early infant production. Extending these findings to the results in this study, it may be that infants of 10-11 months are attending longer to the words that best fit their own production capacities. Both DePaolis and colleagues (2013) and Majorano and colleagues (2014) found evidence for sensitivity between perception and early production at this stage, as infants attended longer to words which contained consonants that were most stable in their babble production (the 'articulatory filter', see Vihman, 1993). In the present study, infants on the verge of production may have been attending longer to those forms that best matched their developing production routines. Infants' reported production ability was low – most infants had not yet produced their first word, and only three infants had five words in their lexicons. As the infants were likely to be on the verge of moving from babble to words, it is possible that their looking behaviour reflects the direction of their developing phonological capacity.

Furthermore, our findings did not show a straightforward advantage for OWs. One might expect that an iconic advantage would show up across all stimuli, regardless of experience of the forms in question. It is also unlikely that difficulty of understanding was a problem for the infants in this experiment. Bergelson and Swingley (2012) showed that infants as young as six months were able to demonstrate understanding of common nouns in an eye-tracking study, and so we would expect that by 10-11 months, if an iconic advantage is present in early word learning, then infants should be able to recognise forms such as *baa* and *vroom*, but this was shown not to be the case. Again, this might support the 'articulatory filter' hypothesis (Vihman 1993), as infants are more

inclined to attend to those forms that match their developing output capacity. However, neither an iconic advantage nor the articulatory filter would preclude the importance of experience in the infants' responses. This goes hand-in-hand with the results from the CDI, which showed *sheep* and *car* to be the least-commonly understood of all the forms.

When aligning our results with the CDI reports, it seems that experience of the stimuli is a key factor. However, infants' responses were not consistent with their own CDI results; instead, an effect was found only when results from the CDI reports were considered on a group level. Parental CDI reports have been previously shown to be accurate for older infants (e.g. Miller et al., 1995), and our results suggest that this is also the case for infants of 10-11 months at the group level, but perhaps less so individually. Of course, reporting infants' word understanding is a more difficult task than reporting their production, and since so few of the infants were able to speak, we must expect some discrepancy between what the infants were reported to understand, and what they could actually understand. Evidence from [blinded for review] (2016) shows that caregivers tend to produce OWs and CWs in equal proportion in the input. This stands up to both the CLEX data (Jørgensen et al., 2010) and the CDI reports, where reported understanding was almost equal across OW and CW forms. We would therefore expect experience of the stimuli to be relatively equal across conditions.

In considering how caregivers judge a word to be 'understood' by their infants, one might suggest that these evaluations are intrinsically linked to the caregivers own production, and whether or not they generally use a specific word, such as *woof woof* or *duck*, in interactions with their infants. The CDI reports may therefore provide an overview of the infants' input experience: the words they are reported to understand may in fact be the words that they have encountered most often in the input. *Sheep* and *car* may in fact be the least commonly *experienced* words in the input, and this would no doubt lead to less reliable form-meaning pairings in an experimental paradigm.

We might expect a true perceptual advantage to generate shorter response latencies in an eye-tracking experiment, and not only longer fixation times. Contrary to this, our response latency analyses generated no interpretable results. However, since the data were substantially reduced for this analysis, to the point where only one third of trials were included, it may simply be a case of an insufficient sample; a more highly-powered replication of this study is required if we are to understand how best to interpret the response latency results. Interestingly, in Figure 2 we note that CWs *car* and *sheep* generated some of the fastest looking times, though again the short fixation times for these stimuli point to a lack of understanding. This leads us to consider the sub-analysis, which showed some interesting trends across the data. The differences shown in fixation proportions across the conditions are convincing, and the grouped CW response, which falls just below chance, is especially noteworthy: while infants are reliably looking towards the target image in the OW condition, they do not match the target image to the CW even 50% of the time. This is consistent across the stimuli when the results are broken down, and suggests that knowledge of the OW stimuli may be better established that that of the CW stimuli.

Another factor which may bias infants' responses towards OWs over CWs is their presentation in the input. All things being equal, if experience of the stimuli was the main point in question, then we should observe no difference between perception of OW and CW forms. Findings from [blinded for review] (2016) present another important advantage for OWs over CWs in early perception: prosodic salience. The authors analysed naturalistic data to show that OWs stand out from the speech stream with regard to pitch and duration, as well as being more easily segmentable due to frequent repetition and reduplication, or presentation in isolation. CWs benefited from none of these advantages. One might posit that, in relation to their OW counterparts, CW acquisition is at a disadvantage in early perception despite the fact that these forms occur with the same frequency. This might explain the low looking in the CW condition in our results. Indeed, there is ample evidence in the literature to show that infants are better able

to learn words which stand out from the input owing to prosodic salience or segmentability (Brent & Siskind, 2011; Fernald & Kuhl, 1987; Floccia et al., 2016), and so we cannot overlook the potential effects of input on our results.

Numerous studies have shown that infants are able to detect sound symbolic properties from a very early age. One drawback of the present study is the age of infants tested: by 10 months, infants have ample linguistic experience, and so it is necessary to test much younger infants if the question of an intrinsic advantage for iconicity is to be fully understood. Furthermore, evidence from the CDI reports showed that the infants had all been previously exposed to the stimuli used in this experiment. This is one advantage of using non-words in experimental research, as this assures no previous exposure of the forms in question. Of course, the disadvantage of such studies is that they are largely inapplicable to the reality of language learning. Testing English-learning infants on novel mimetic or sound-symbolic words does not reflect the reality of these infants' language-learning processes, as we could argue that infants do not always have the potential advantage of mimetics in the ambient language to draw upon during acquisition. In testing onomatopoeia – words which are found across many (if not all) of the world's languages – we have attempted to determine whether the advantages suggested with regard to language learning and iconicity may be pertinent in real-time language learning. However, it is impossible to separate the potential iconic effect that OWs may possess with their other attributes: the fact that they are more salient in the early input ([blinded for review], 2016) and typically suited to the early output owing to their simple phonological forms ([blinded for review], 2014) may also have been driving infants' responses. There is no doubt that a multitude of factors contribute to infants' learning of one lexical item over another, and in the case of onomatopoeia this is made even more complex by the question of iconicity. While we cannot exclude the possibility of an iconic advantage from infants' perception of onomatopoeia, it should be acknowledged that this cannot be separated from infants' experience of these forms in both the input and the output. In

order to test the role of onomatopoeia more thoroughly, a highly controlled experiment would be required, with learning of novel OW-like forms in contrast with CW-like forms. In reality, owing to the nature of onomatopoeia as 'sound effect' words, early experience of these forms is likely to be engaging with regard to prosody, semantics and phonology. This is a feature of their iconic status in language, but also a potential confound when attempting to question the notion of an iconic advantage for these forms.

Conclusion

Our results suggest a perceptual advantage for onomatopoeia over their conventional counterparts in early language development. However, results were not consistent across the stimuli, and appeared to be driven at least in part by linguistic experience. Infants showed significant recognition of some OW targets, while their recognition of CWs was unreliable. Results did not generate longer looking in OWs across-the-board, but only in those forms that were most commonly understood by the infants. While these results can be interpreted as suggesting some iconic advantage for onomatopoeia, we propose that their status in the input – with regard to prosody, production and semantics – is also an important factor that cannot be considered independently of iconicity.

Acknowledgements

The author would like to thank Marilyn Vihman, Tamar Keren-Portnoy and Elika Bergelson for feedback on this paper, as well as Ben Miller and Tim Holmes for their assistance with programming and Jessica Wormald for providing the audio stimuli. This research was funded by the Economic and Social Research Council, grant number ES/J500215/1, and was originally submitted as part of the author's doctoral thesis.

References

Asano, M., Imai, M., Kita, S., Kitajo, K., Okada, H. & Thierry, G. (2015). Sound symbolism scaffolds language development in preverbal infants. *Cortex* **63**, 196-205.

Behrens, H. (2006). The input–output relationship in first language acquisition. *Language and Cognitive Processes* **21**, 2-24.

Bergelson, E. & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences* **109**, 3253-3258.

Bohn, M., Call, J., & Tomasello, M. (2016). Comprehension of iconic gestures by chimpanzees and human children. *Journal of Experimental Child Psychology* **142**, 1-17.

Brent, M. R. & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition* **b**, B33-B44.

Canfield, R. L., Smith, E. G., Brezsnyak, M. P., Snow, K. L., Aslin, R. N., Haith, M. M., Wass, T. S. & Adler, S. A. (1997). Information processing through the first year of life: A longitudinal study using the visual expectation paradigm. *Monographs of the Society for Research in Child Development* **62**, i-160.

Davis, R. (1961). The fitness of names to drawings: A cross-cultural study in Tanganyika. *British Journal of Psychology* **52**, 259-268.

DePaolis, R. A., Vihman, M. M. & Nakai, S. (2013). The influence of babbling patterns on the processing of speech. *Infant Behaviour and Development* **36**, 642-649.

Dingemanse, M. (2012). Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass* **6**, 654-672.

D'Onofrio, A. (2014). Phonetic detail and dimensionality in sound-shape correspondences: Refining the bouba-kiki paradigm. *Language and Speech* **57**, 367-393.

Elsen, H. (1991). *Erstspracherwerb: Der Erwerb des deutschen Lautsystems*. Wiesbaden: Deutscher Universitaets Verlag.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., ... & Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development* **59**, i-185.

Fernald, A. & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behaviour and Development* **10**, 279-293.

Fernald, A., Pinto, J. P., Swingley, D., Weinbergy, A. & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science* **9**, 228-231.

Fernald, A., Zangl, R., Portillo, A. L. & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language. In I. A. Sekerina, E. M. Fernández & H. Clahsen (eds.) *Developmental Psycholinguistics: On-line methods in children's language processing* (pp.97-136). Amsterdam: John Benjamins.

Fikkert, P. & Levelt, C. (2008). How does place fall into place? The lexicon and emergent constraints in children's developing phonological grammar. In P. Avery, E. Dresher & K. Rice (eds.) *Contrast in phonology: theory, perception and acquisition* (pp. 231-270). Berlin: Mouton.

Floccia, C., Keren-Portnoy, T., DePaolis, R., Duffy, H., Delle Luche, C., Durrant, S., & Vihman, M. (2016). British English infants segment words only with exaggerated infant-directed speech stimuli. *Cognition* **148**, 1-9.

Fort, M., Weiß, A., Martin, A. & Peperkamp, S. (2013). Looking for the bouba-kiki effect in prelexical infants. *Proceedings of the 12th International Conference on Auditory–Visual Speech Processing, [Annecy, France, 29 August–1 September 2013],* 71-76.

Hamilton, A., Plunkett, K. & Schafer, G. (2000). Infant vocabulary development assessed with a British communicative development inventory. *Journal of Child Language* **27**, 689-705.

Hinton, L., Nichols, J. & Ohala, J. J. (1994). Introduction: Sound symbolic processes. In L. Hinton, J. Nichols & J. Ohala, (eds.) *Sound symbolism* (pp.1-12). Cambridge, UK: Cambridge University Press.

Imai, M. & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**, 20130298.

Imai, M., Kita, S., Nagumo, M. & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition* **109**, 54-65.

Ivanova, G. 2006. Sound-symbolic approach to Japanese mimetic words. *Toronto Working Papers in Linguistics* **26**, 103-114.

Jørgensen RN, Dale PS, Bleses D, Fenson L. (2010). CLEX: A cross-linguistic lexical norms database. *Journal of Child Language* **37,** 419–428.

Kita, S. (1997). Two-dimensional semantic analysis of Japanese mimetics. *Linguistics* **35**, 379-416.

Köhler, W. (1970). *Gestalt psychology: An introduction to new concepts in modern psychology* (9th ed.). New York, NY: New American Library.

Kunnari, S. (2002). Word length in syllables: Evidence from early word production in Finnish. *First Language* **22**, 119-135.

[blinded for review] (2014). A phonological analysis of onomatopoeia in early word production. *First Language* **34**, 387-405.

[blinded for review] (2016). How salient are onomatopoeia in the early input? A prosodic analysis of infant-directed speech. *Journal of Child Language.* [FirstView Article]

Lockwood, G., Dingemanse, M., & Hagoort, P. (2016). Sound-symbolism boosts novel word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **42**, 1274-1281.

Lyons, J. (1968). *Introduction to theoretical linguistics*. Cambridge, UK: Cambridge University Press.

Majorano, M., Vihman, M. M. & DePaolis, R. A. (2014). The relationship between infants' production experience and their processing of speech. *Language Learning and Development* **10**, 179-204.

Maurer, D., Pathman, T. & Mondloch, C. J. (2006). The shape of boubas: sound–shape correspondences in toddlers and adults. *Developmental Science* **9**, 316-322.

Menn, L. & Vihman, M. M. (2011). Features in child phonology: inherent, emergent, or artefacts of analysis? In N. Clements & R. Ridouane (eds.), *Where Do Phonological Features Come From? The nature and sources of phonological primitives.*(Language Faculty and Beyond 6.) Amsterdam: John Benjamins.

Miller, J. F., Sedey, A. L., & Miolo, G. (1995). Validity of parent report measures of vocabulary development for children with Down syndrome. *Journal of Speech, Language, and Hearing Research* **38**, 1037-1044.

Monaghan, P., Shillcock, R.C., Christiansen, M.H. & Kirby, S. (2014). How arbitrary is language? *Philosophical Transactions of the Royal Society B* **369**, 20130299.

Nygaard, L. C., Cook, A. E. & Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition* **112**, 181-186.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica* **41**, 1-16.

Ozturk, O., Krehm, M. & Vouloumanos, A. (2013). Sound symbolism in infancy: evidence for sound–shape cross-modal correspondences in 4-month-olds. *Journal of Experimental Child Psychology* **114**, 173-186.

Peirce, J. W. (2007). PsychoPy - psychophysics software in Python. *Journal of Neuroscience Methods* **162**, 8-13.

Perry, L. K., Perlman, M., & Lupyan, G. (2015). Iconicity in English and Spanish and its relation to lexical category and age of acquisition. *PloS one* **10**, e0137147.

R Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/.

Sapir, E. (1970). *Language: An introduction to the study of speech*. London, UK: Rupert Hart-Davis.

Sasamoto, R., & Jackson, R. (2016). Onomatopoeia–Showing-word or Saying-word? Relevance Theory, lexis, and the communication of impressions. *Lingua* **175**, 36-53.

Saussure, F. de (1962). *Cours de linguistique générale*. 5th Edition. Paris, France: Payot.

Swingley, D. & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition* **76**, 147-166.

Swingley, D., Pinto, J. P. & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition* **71**, 73-108.

Tardif, T., Fletcher, P., Liang, W., Zhang, Z., Kaciroti, N. & Marchman, V. A. (2008). Baby's first 10 words. *Developmental Psychology* **44**, 929-938.

Vihman, M. M. (1993). Variable paths to early word production. *Journal of Phonetics* **21**, 61-82.

Vihman, M. (2016). Prosodic structures and templates in bilingual phonological development. *Bilingualism: Language and cognition* **19**, 69-88.

Vinson, D. P., Cormier, K., Denmark, T., Schembri, A., & Vigliocco, G. (2008). The British Sign Language (BSL) norms for age of acquisition, familiarity, and iconicity. *Behavior Research Methods* **40**, 1079-1087.

Werner, H. & Kaplan, B. 1963. *Symbol formation*. New York, NY: John Wiley & Sons.