Word frequency, function words and the second gavagai problem

Jean-Rémy Hochmann *  
SISSA, International School for Advanced Studies, Cognitive Neuroscience Department, via Bonomea, 265, 34136 Trieste, Italy  
Harvard University, Department of Psychology, William James Hall, 33 Kirkland Street, Cambridge, MA 02138, USA

Abstract

The classic gavagai problem exemplifies the difficulty to identify the referent of a novel word uttered in a foreign language. Here, we consider the reverse problem: identifying the referential part of a label. Assuming “gavagai” indicates a rabbit in a foreign language, it may very well mean “a rabbit” or “that rabbit”. How can a learner know whether rabbit is actually said “vagai”, “gava” or “gavagai”? Here, we report evidence suggesting that infants can identify potential function words on the basis of their high frequency and avoid considering them when associating labels and referents. In three experiments, 17-month-old infants were first exposed to an artificial speech stream where frequent and infrequent syllables alternated (e.g., ... gibuvokugifevodegita...). Infants then saw a novel object and heard the repetition of a bisyllabic label consisting of one frequent and one infrequent syllable (e.g., vomu). The frequent syllable was the initial syllable of the label in Experiment 1 and the final in Experiments 2 and 3. We then presented infants with both the previous and now familiar object and a novel object. We asked whether infants would be more likely to orient first towards the familiar object when hearing a label with a new frequent and the previous infrequent syllables (e.g., gimu), or when hearing a label with a new infrequent and the previous frequent syllables (e.g., vona). Results suggest that the infrequent syllable was associated more strongly with the object, than the frequent one, only when the perceived position of the frequent syllable was constant all along the experiment.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

1.1. The second gavagai problem

Quine (1960) famously introduced the “gavagai” problem, whereby a child acquiring language is facing the problem of the indeterminacy of translation when trying to understand the meaning of a novel word. In Quine’s metaphor, children are in the same situation as an adventurer witnessing an island native that would point at a running rabbit and say “gavagai”. Does “gavagai” refer to the rabbit, its ears, its color, its running or something else? This problem has since received several solutions, involving the use of conceptual biases (Markman, 1990; Markman, Wasow, & Hansen, 2003), socio-pragmatic (Akhtar & Tomasello, 1996; Baldwin, 1991, 1993; Csibra, 2003; Hirsh-Pasek, Golinkoff, & Hollich, 2000; Sperber & Wilson, 2004) and syntactic cues (Bernal, Lidz, Millote, & Christophe, 2007; Brown, 1957; Gleitman, 1994; Naigles & Kako, 1993; Waxman & Booth, 2001, 2003). However, Quine’s situation contains a second problem that has hardly been addressed. Assuming that the learner solved the classic gavagai problem and identified the referent, what part of speech actually refers to the referent? In other words, given that the island native intended to name the rabbit, is rabbit “gavagai” in his language? Or is it “gava”, “vagai” or “gai”? This second gavagai problem is not anecdotal, as words, especially nouns, are rarely pronounced in isolation. Only 7–12% of child directed speech utterances consist of
isolated words (Brent & Siskind, 2001; Christiansen, Allen, & Seidenberg, 1998; Fernald & Morikawa, 1993; Fernald & Simon, 1984). In particular, nouns in languages like English or Italian are usually associated with a determiner. Therefore, instead of just “rabbit”, “gavagai” may well mean “a rabbit”. Learners would thus need to strip the part that corresponds to the determiner, before pairing the noun and its referent.

A solution to the second gavagai problem may involve early knowledge about a category of determiners, or more broadly of function words. If infant learners discriminate between function and content words and know that function words are not referential, they may correctly focus on content words when learning the label of a novel object or kind.

1.2. Early representations of syntactic categories

The question of whether young children and infants represent syntactic categories has been debated in the field of language acquisition. One view proposes that young children rely on item-based templates that are syntactically unanalyzed (Dabrowska, 2001; Tomasello, 2003). According to such item-based or usage-based theory of language acquisition, children initially store a certain amount of fixed sentence templates. This view predicts that children’s initial production merely reflects their input and is not productive. A child that says “eat pasta” and “want cookie” would not necessarily say or understand “eat cookie” and “want pasta”. Support to this theory mainly comes from corpus analysis of infants’ spontaneous productions (Pine & Martindale, 1996; Tomasello, 2003).

This view is contradicted, however, by studies showing productive use of morphological and syntactic constructions, even leading to overgeneralization (Marcus et al., 1992). For example, in a classic study, Berko (1958) introduced 4-year-old preschool children with a novel animal that she called a wug: “this is a wug”. She then showed two exemplars of the same animal, and asked children to complete her sentence: “these are two...”. Children spontaneously answered “wugs!”, generalizing the rule for plural formation in English to the novel word they had just learned in the singular form. Comprehension studies suggest that this ability initially develops between 20- and 24-months of age, well before production, and is further refined in the third year after birth (Kouider, Halberda, Wood & Carey, 2006).

Moreover, Lidz, Gleitman, and Gleitman (2003) showed that 3-year-olds already analyze the structure of simple sentences, and use that structure to constrain their interpretation of the sentence (see also Lidz & Gleitman, 2004). They studied learners of Kannada, a language where causative meaning can be marked in two ways: by a morphological affix on the verb or by a transitive construction, as it is classically done in English. In infants’ and children’s input, causative morphological marking is more frequently and more consistently associated with causative meaning than is the transitive structure. The usage-based theory would thus predict that young children should be better at interpreting the morphological marker construction than the transitive construction. Lidz et al. (2003) showed that this is not the case. Three-year-old Kannada learners based their interpretation of the meaning of a novel verb on the number of arguments (one or two) rather than on the presence or absence of the causative morphological marker. Thus, rather than using the meaning of template sentences to learn about the structure of their language, Kannada learners rely on a structural analysis to learn the meaning of novel verbs. In that respect, they behave as young learners of English (Arunachalam & Waxman, 2010; Naigles & Kako, 1993; Yuan & Fisher, 2009), Mandarin (Lee & Naigles, 2005, 2008), or Turkish (Gökşun, Küntay, & Naigles, 2008).

Beyond lexical categories such as nouns and verbs, the debate has also revolved around the question of whether young children have functional categories such as determiners. In particular, researchers have asked whether young children show overlap in their use of determiners. For example, do infants who have learned “a ball” and “the duck” also say “the ball” and “a duck”? Or is their production item-based? Analyzing children’s production corpora and using different criteria, Pine and colleagues (Pine & Lieven, 1997; Pine & Martindale, 1996) argued for the latter, while Valian and colleagues (Valian, 1986; Valian, Solt, & Stewart, 2009) argued for the former possibility. Yang (2010) recently proposed that the pattern of children’s production could be explained by the general phenomenon that relatively few words are used very frequently, whereas most words are rarely used (a statistical distribution known as Zipf’s law). Reanalyzing production data taking this distribution into account, Yang (2010) claimed that the data ultimately support the view that very young children are already equipped with productive syntactic rules.

Other evidence for the representation of syntactic functional categories consists in how these constrain word-to-world mapping. For example, Waxman and Booth (2003) showed that 14-month-old infants associate a novel word to an object kind if it is preceded by a determiner (the blic-ket) and to a property of the object if ending in -ish and followed by the pronoun one (the blickish one). Bernal et al. (2007) showed that 23-month-old infants map a novel word onto a novel action if this word appeared following a pronoun, but not if it followed a determiner (see also Brown, 1957).

In our opinion, experimental evidence thus favors the view that young children and infants as young as 14-months have syntactic categories, or at least precursors of these categories. How do they acquire or recognize these categories? Distributional analysis and phonological cues may play a role in this process.

1.3. Forming syntactic categories

A relatively successful strategy to classify words into nouns and verbs is to classify them according to the frames

---

1 The second gavagai problem, we have to stress, cannot be reduced to the problem of segmentation, which has been the focus of much investigation in recent years (Aslin, Saffran, & Newport, 1998; Bonatti, Peña, Nespor, & Mehler, 2005; Buiatti, Peña, & Dehaene-Lambertz, 2009; Gómez, Bion, & Mehler, 2011; Mehler, Peña, Nespor, & Bonatti, 2006; Safran, Aslin, & Newport, 1996; Shukla, Nespor, & Mehler, 2007 and many others). Even after one has segmented gavagai into distinct words such as ga and vagai, one still needs to identify which of these words refers to rabbit.
in which they occur, i.e., the sets of words that precede and follow them (Mintz, 2003; Mintz, Newport, & Bever, 2002). This strategy is not just an abstract possibility: behavioral experiments suggest that adults can use such a mechanism for classifying non-sense words into artificial categories (Mintz, 2002). Furthermore, this mechanism is already active in infants. Mintz (2006) showed that 12-month-olds can use frequent frames made of function words to categorize novel words. Infants were introduced to two novel words in typical noun frames (e.g. “I see the gorp in the room”) and two novel words in typical verb frames (e.g. “She wants to lonk it”). Infants next listened to all the novel words in both noun and verb frames. They showed an increased looking time for novel nouns presented in a verb frame (e.g. “I gorp you now!”), but not for novel verbs presented in a noun frame (e.g. “Here’s a lonk of a dog”). This suggests that infants understood that the novel nouns could appear only in certain frames, thus forming a distributional category. The classification of verbs, however, was not yet effective with the used frames. With a similar method, Höhle et al. (2004) found that German 14- to 16-month-olds but not 12- to 13-month-olds discriminated passages containing a novel word in noun contexts from passages containing that same word in verb contexts; this occurred if the novel word (e.g., “glamm”) had been previously introduced to the child preceded by a deter- mine (e.g., “ein glamm”), but not if it had been introduced preceded by a pronoun (e.g., “sieg glamm”). They concluded that infants of this age can use a frequent determiner to classify a novel word (e.g., “sie glamm”), but not if it had been introduced preceded by a function word (e.g., “She wants to lonk it”). Infants next listened to all the novel words in both noun and verb frames. They showed an increased looking time for novel nouns presented in a verb frame (e.g. “I gorp you now!”), but not for novel verbs presented in a noun frame (e.g. “Here’s a lonk of a dog”). This suggests that infants understood that the novel nouns could appear only in certain frames, thus forming a distributional category. The classification of verbs, however, was not yet effective with the used frames. With a similar method, Höhle et al. (2004) found that German 14- to 16-month-olds but not 12- to 13-month-olds discriminated passages containing a novel word in noun contexts from passages containing that same word in verb contexts; this occurred if the novel word (e.g., “glamm”) had been previously introduced to the child preceded by a determiner (e.g., “ein glamm”), but not if it had been introduced preceded by a pronoun (e.g., “sieg glamm”). They concluded that infants of this age can use a frequent determiner to classify a novel word as a noun, but the ability to use pronouns to categorize novel verbs might still be immature (see also Shi & Melançon, 2010 for similar results with French-learning 14-month-olds). Thus, infants appear capable of using some function words to form distributional categories at the beginning of the second year after birth.

Other authors insisted on the role of correlations between syntactic category membership and phonological properties such as stress pattern, syllabic structure and vowel reduction (Monaghan, Chater, & Christiansen, 2005). Monaghan, Christiansen, and Chater, (2007) further showed that the combination of distributional cues comparable to the frames discussed above, and phonological cues constituted a good predictor of syntactic category membership. Their approach showed the plausibility of category formation on the basis of perceptual and distributional cues, exemplifying the actual richness of the input provided to infant learners. However, the phonological cues that are actually useful vary across languages. To give only one example, the number of syllables is useful to discriminate nouns from verbs in English but not in Dutch, French or Japanese (Monaghan, Christiansen, & Chater, 2007). Infants must therefore first identify what specific perceptual cues play a role in their language. Even then, we must stress that there remains a large overlap between different syntactic categories along any phonological dimension (Monaghan, Chater, & Christiansen, 2005). Therefore, the combination of multiple perceptual and distributional cues appears necessary for accurate categorization (Monaghan et al., 2007).

These studies, however, show only evidence of the formation of categories based on distributional and perceptual properties; they do not test for the labelling of those categories (Mintz, 2006). Indeed, they show that in certain frames, only certain words can appear, but the labeling of a category as ‘nouns’ or ‘verbs’, or ‘function words’ or ‘content words’, should come from other sources of information. In fact, labeling distributional categories represents a fundamental problem of language acquisition, for which solutions are likely to necessitate prior linguistic knowledge. For example, Pinker (1984) proposed that infants are endowed with the knowledge of the syntactic categories ‘nouns’ and ‘verbs’ as well as the rules to map these categories on the world.

1.4. Identifying function words

As mentioned above, a solution to the second gavagai problem may involve identifying or forming a category of function words. Infants start recognizing some function words in the second semester after birth. For example, 11- to 13-month-old infants discriminate passages containing real function words from passages where the function words were replaced by non-words, both in behavioral studies (Shi, Werker, & Cutler, 2006) and in electrophysiological studies (Shafer, Shucard, Shucard, & Gerken, 1998). By 8 months, English-learning infants can recognize the most frequent function word (i.e., the) in their language but do not represent it in full phonetic details (Shi, Cutler, Werker & Cruickshank, 2006). Indeed, they recognize and segment from a phrase the word the but also the mispronounced form kuh. By 11-months, English-learning infants recognize only the correct form the. A similar picture was observed for French-learning infants, but at a younger age (Shi & Lepage, 2008; Shi, Marquis, & Gauthier, 2006). These studies, however, do not mean that infants have identified these words as function words, but solely that they recognize the most frequent elements in their language.

Both phonological and distributional cues may help infants to form or identify a category of function words (Cutler, 1993; Hochmann, Endress, & Mehler, 2010; Shi, Morgan, & Allopenna, 1998; Shi, Werker, & Morgan, 1999). Compared to content words, function words are usually shorter, simpler and unstressed, properties to which even newborns are sensitive (Shi et al., 1999). In addition, function words are the most frequent words in a language (Gervain, Nespor, Mazuka, Horie, & Mehler, 2008; Shi et al., 1998).

High frequency of occurrence has been linked to particular uses. Artificial grammar learning studies have long suggested that frequent elements play an important role in the acquisition of positional rules (Braine, 1963, 1966; Green, 1979; Valian & Coulson, 1988; Valian & Levitt, 1996). In particular, frequent elements may act as anchor
Recent studies further suggest that infants use frequent words to form a category of words. Exposing 7-month-old infants to an artificial language, Gervain et al. (2008) showed that infants expect very frequent words to occupy the same position that frequent (function) words most consistently occupy in the language of their environment (i.e., initial for Italian learners and final for Japanese learners). Thus, this suggests that frequent words form a category and that a distributional property observed for members of this category is extended to new members. Is this category solely a byproduct of infants’ sensitivity to distributional properties, or do they use it in a specific way that may help them construct or identify the category of function words?

In a recent paper, we showed that infants tend to associate an object referent with an infrequent stressed word, rather than with a very frequent unstressed word (Hochmann et al., 2010). In that study, Italian infants were first habituated to a series of sentences in a language unknown to them (i.e., French). The French sentences were constructed so that two determiners appeared nine times more frequently than any other word. One of these determiners, and a noun were then paired to a novel object. We showed that infants formed stronger associations between the object and the noun when the object and the determiner did not differ in pitch and duration. In three experiments, we asked whether infants are more likely to associate an object with a frequent or an infrequent syllable, when these do not differ in pitch and duration. In each experiment, we first exposed infants to an artificial speech stream where one of two frequent syllables alternate with one of eighteen relatively infrequent syllables. In a second phase, infants saw an object on the screen and heard the repetition of a bisyllabic label consisting of one frequent and one infrequent syllable. We then presented infants with both the previous and now familiar object and a novel object. We asked whether infants would be more likely to orient first towards the familiar object when hearing a label with a new frequent and the previous infrequent syllables, or when hearing a label with a new infrequent and the previous frequent syllables. In other terms, do they form a stronger association between the object and the frequent syllable or between the object and the infrequent syllable? The three experiments differed only for the initial exposure phase. Experiments 1 and 2 were identical, except for the syllables that were frequent in the familiarization. In Experiment 1, the frequent syllables were the first syllables of the object labels. In Experiment 2, the frequent syllables were the final syllables of the object labels. Finally, Experiment 3 was identical to Experiment 2, except that the familiarization stream was segmented (see Fig. 1).

### 2. Experiment 1

#### 2.1. Materials and methods

Fig. 1 presents the experimental paradigm for Experiments 1, 2, and 3. Each experiment consisted of three phases: a familiarization phase, a teaching phase, and a test phase. Two successive phases were separated by a visual fixation attractor, that is, a white cross presented centrally on the screen and moving back and forth. The experimenter started each phase by pressing a key when the infant was looking at the central fixation attractor.

#### 2.1.1. Participants

Forty-three Italian 17-month-old infants were tested. Twenty were excluded for fussiness (7), equipment failure (insufficient eye tracker data, 8), the mother not following experimental instructions (2) or side bias (3). The remaining 23 infants (16 males, 7 females, age range: 17 month and 4 days–18 month and 6 days) were included in the final analysis.

#### 2.1.2. Stimuli

##### 2.1.2.1. Familiarization stream

In the familiarization phase, we presented infants with an artificial speech stream. The stream alternated between one of two very frequent syllables (gi or vo) and one of eighteen infrequent syllables. Nine syllables could follow the frequent syllable gi (bu,
ko, fe, da, ni, pu, si, ta, go) and nine different syllables could follow the frequent syllable vo (bo, ku, fi, mu, na, pi, de, ga, va). While the two frequent syllables alternated, the infrequent syllables were homogeneously distributed all over the stream: no infrequent syllable could appear for the \((n + 1)\)th time before all the other infrequent syllables had appeared \(n\) times. Each frequent syllable appeared 126 times and each infrequent syllable appeared 14 times. The two frequent syllables were therefore nine times more frequent than each of the infrequent syllables.

The stream was generated by the speech synthesizer MBROLA (fr4) with a phoneme duration of 120 ms and a monotonous pitch of 200 Hz. Frequent and infrequent syllables did not differ in terms of pitch and duration. The beginning and the end of the stream were ramped in amplitude. The whole familiarization stream lasted 2 min.

### 2.1.2.2. Objects.

The objects used in the teaching and test phases were simple three-dimensional shapes generated as 3D animations in Maya 6.0 (Autodesk, Inc., San Rafael, CA), using a frame rate of 25 fps, the H.264 codec and the mov container format. The two objects are visible in Fig. 1. One was a blue three-dimensional cross. The other was a green pile of rings (similar to the belly of the emblem of the Michelin tire brand). Both objects were symmetrical, had similar perceived volumes and perimeters, and were found to be equally attractive for 17-month old infants in a pilot study. However, the two objects had clearly different shapes and colors. During the teaching phase, one of the objects was presented in an animated movie of 33 s. During this movie, the object moved from one side to the other while rotating around its axes.

![Fig. 1. Schematic representation of the design of Experiments 1, 2 and 3. See the main text for more details.](image)

### 2.1.2.3. Labels.

Four bisyllabic labels were generated, associating the two frequent syllables (gi and vo) and two infrequent syllables of the familiarization (mu and na), thus generating the labels vomu, gina, vona and gimu. Two of these labels were used in the teaching phase: vomu or gina. The other two labels were used in the test phase: gimu and vona. The labels were generated by the speech synthesizer MBROLA (fr4) with a phoneme duration of 120 ms and a monotonous pitch of 200 Hz.

### 2.1.3. Apparatus

In this and all other experiments reported here, infants were tested individually. They sat on a parent’s lap 80 cm from a 17-inch LCD screen in a dimly lit, sound-attenuated cubicle. Parents wore dark sunglasses throughout the experiment to avoid all parental influence on the infants’ behavior.

### 2.1.4. Procedure

#### 2.1.4.1. Familiarization.

The familiarization stream was played from a loudspeaker located behind the screen.
2.1.4.2. Teaching phase. In the teaching phase, infants saw a visual object on the screen and simultaneously heard a speech label. We used the two shapes described above and two labels composed of a frequent and an infrequent syllable from the familiarization stream. The four stimulus combinations resulting from two objects and two labels were counterbalanced across infants. Each infant learned only one object-label combination.

During the teaching phase, the object moved from one side to the other on an LCD screen while rotating around its axes. Simultaneously, a label was repeated 24 times. Two repetitions of the label were separated by 900 ms of silence, yielding a total duration of 32 s.

The presentation movie could be interrupted if the infant looked away for more than 2 s. However, all the 23 infants included in the analysis looked at the entire movie without interruption.

2.1.4.3. Test phase. Following the familiarization phase and the teaching phase, infants completed four test trials. The trials were separated by the presentation of the central fixation attractor (i.e. a white cross moving back and forth). The experimenter started each trial by pressing a key when the infant was looking at the central fixation attractor.

Once the experimenter started a trial, the fixation attractor disappeared, and the two objects visible in Fig. 1 appeared on the computer screen. Each of the objects on the screen occupied an area of 8.5 cm × 9 cm. The centers of the two objects were separated by about 19 cm. For each infant, the position of objects was counterbalanced across trials. We also counterbalanced the position in which the blue object appeared in the first trial across infants. The objects remained static and visible on the screen for the entire trial duration. Two seconds after the trial started, the central fixation attractor appeared again for 3.32 s. A test label (see below) was first pronounced while the central fixation attractor was still visible on the screen; the offset of the label was synchronized with the attractor’s disappearance. Following this, the test label was repeated four more times, two consecutive repetitions being separated by 900 ms of silence. The two objects remained visible on the screen for 7 s after the central fixation disappeared, until the end of the trial. As infants were allowed to see the position of the two objects on the screen before a test label was pronounced, we expected them to look first and longer at the object they thought was being named.

Each infant completed two test trials in the “same-frequent” condition, and two test trials in the “same-infrequent” condition. Two consecutive trials were of different conditions. The two conditions were identical except for the test label used. In both conditions, the label differed only in one syllable from the label heard during the teaching phase. In the same-frequent condition, the label had the same frequent syllable as during the teaching phase, but a different infrequent syllable. In the same-infrequent condition, the label had the same infrequent syllable as during the teaching phase, but a different frequent syllable (see Fig. 1).

2.1.5. Analysis

We defined two areas of interest for the analysis of the infant eye gaze. Each was a rectangle of 11 cm × 13 cm, centered on one of the objects. Infants’ looking behavior was monitored using a Tobii 1750 Eye-tracker system and the Clearview 2.5.1 software package. Infants’ gaze position was monitored with a frequency of 50 Hz (i.e. one data point every 20 ms). Clearview was used to acquire and export the gaze information in terms of fixations defined by a radius of 30 pixels and minimum duration of 100 ms. Only infants for whom the eye tracker data for each trial contained at least one fixation of at least 100 ms in one of the two areas of interest were included in the analysis. Infants not meeting these criteria were excluded for insufficient eye tracker data. Considering the four test trials, infants for whom more than 70% of the total fixation time was spent in only one of the areas of interest were considered to exhibit a side bias and were rejected from further analysis.

2.1.5.1. First look analysis. As a dependent variable in this analysis, we considered the first look to one of the objects following the first presentation of the test label. The first look was defined as the first uninterrupted fixation of at least 100 ms in one of the two areas of interest described above, after the central attractor disappeared and before the end of the trial.

In each test trial, infants were coded either as first looking at the familiar object from the teaching phase, or as first looking at the novel object. In each trial, infants scored 1 if they first looked at the familiar object and 0 if they first looked at the novel object. With two trials per condition, each infant could obtain a total score of 0, 1 or 2 in each condition. We then assessed whether these numbers differed between the conditions using a two-tailed Wilcoxon signed rank test.

2.1.5.2. Preferential looking analysis. Word learning experiments traditionally rely on preferential looking paradigms based on measures of cumulative looking time by the end of each trial. Here, profiting from the amount of data provided by the Tobii Eyetracker, we present an exploratory analysis showing the dynamics of preferential looking along the time course of each trial.

For each trial, we evaluated infants’ preferential looking for five periods 0–1380 ms, 1380–2760 ms, 2760–4140 ms, 4140–5520 ms and 5520–6900 ms. The first four periods were delimited by the end of two successive utterances of the test label. The fifth period began with the end of the last repetition of the label and lasts 1380 ms. For each period, we computed the proportion of time spent looking at the familiar object. Precisely, for each period, we summed the amount of time spent looking at either object and computed the proportion of that time spent looking at the familiar object: the time spent looking at the familiar object divided by the time spent looking at either object.
For each period, we averaged separately the trials corresponding to the same-frequent condition, and those corresponding to the same-infrequent condition.

For statistical evaluation of the observed effects, we performed a repeated-measures ANOVA with same-syllable (frequent or infrequent) as within-subject factor, and time as a repeated variable. For each time period, we also ran a student’s t-test to compare the proportion of time spent looking at the familiar objects in the same-infrequent and same-frequent conditions.

2.2. Results

2.2.1. First look analysis

First fixations occurred on average 958 ms after the central attractor disappeared (SD = 908 ms). Fig. 2 and Table 1 show how often infants first looked towards the familiar object from the teaching phase as a function of the experimental condition. Infants did not look first significantly more often at either object in the same-infrequent (M = 1.30, SD = .70, Wilcoxon signed rank test, W = 21, P = .092) or in the same frequent condition (M = .87, SD = .46, W = 3, P = .375). Importantly, though, infants were significantly more likely to look first at the familiar object in the same-infrequent condition than in the same-frequent condition, Wilcoxon signed rank test, W = 19.5, P = .037.

To make sure that each infant’s fixation pattern was not based on what they happened to fixate before the offset of the label, we also performed separate analyses for trials where infants already fixated an object (rather than the central attractor) before the offset of the label. In total, there were 17 such trials (18.5% of the trials). Among these 17 trials, eight occurred in the same-infrequent condition and nine in the same-frequent condition. After removing these 17 trials, our central pattern of results remained unchanged: infants were significantly more likely to look first to the familiar object in the same-infrequent condition (M = 1.09, SD = .73) than in the same-frequent condition (M = .57, SD = .51), Wilcoxon signed rank test, W = 19.5, P = .018.

2.2.2. Preferential looking analysis

Overall, the preferential looking analysis, presented in Fig. 3, corroborated the results of the first look analysis. We performed Student’s t-tests to compare the proportion of time spent looking at the familiar object in each of the five time periods of interest against chance (.5), in the same-infrequent and in the same-frequent conditions. Results showed that infants did not significantly differ from chance in either condition in any period; Ps > .6 after Bonferroni corrections for ten comparisons.

We ran a repeated-measures ANOVA with same-syllable (frequent or infrequent) as a within-subject factor and time as a repeated variable. Two participants had no data for one of the time periods in one condition, so that 21 infants were included in this analysis. It revealed a main effect of same-syllable: infants spent more time looking at the familiar object in the same-infrequent condition than in the same-frequent condition; F(1,20) = 5.39; P = .031. There was no significant main effect of time and no significant interaction between time and same-syllable; Ps < 1.

Finally, planned comparisons asked whether infants looked more at the familiar object in the same-infrequent than in the same-frequent condition in each of the five time periods of interest. The difference was significant for the third time period; t(122) = 2.26; P = .034. Other comparisons showed no statistically significant differences; all Ps > .1.

2.3. Discussion

The results of Experiment 1 replicate and extend those previously reported by Hochmann et al. (2010). After listening to a speech stream alternating between one of two very frequent and one of eighteen relatively infrequent syllables, infants paired a novel object more strongly with the infrequent than with the frequent syllable of a label. In contrast to our previous work, the familiarization speech stream was not produced by a natural speaker of a foreign language, but was generated by an artificial speech synthesizer. This allowed us to ensure that all syllables over the experiment had the same pitch and duration. Therefore, the present results suggest that acoustic factors are not necessary to yield the effect first reported by Hochmann et al. (2010) and replicated here. This strengthens our hypothesis that infants consider very frequent syllables less likely to be associated with referents, than infrequent syllables. It is often assumed that a coalition of cues allows infants to recognize the category membership of words in natural languages (Shi et al., 1998). However, our results suggest that distributional cues such as frequency may be sufficient for infants to recognize potential function words.

Nevertheless, in addition to frequency, a second distributional cue may play a role in the effect that we observed.
Indeed, in Experiment 1, as well as in the experiments reported by Hochmann et al. (2010), frequent syllables always appeared in initial position. This position is congruent with the position of determiners and, more generally, of function words in Italian, the language of our participants. If the use of frequent syllables in Experiment 1 reflects a property that infants learned from exposure to their language, the effect might be restricted to initial frequent syllables. In contrast, if the non-referential use of frequent syllables reflects a mechanism supporting early steps of language acquisition, it should be generalizable to final frequent syllables.

In Experiment 2, we asked whether the effect observed in Experiment 1 for initial frequent syllables could be obtained for final frequent syllables. Furthermore, Experiment 2 served as a control for an alternative explanation of our results: infants may simply rely on final rather than initial syllables when learning a novel object label. If this were true, frequency would not play any role in the results of Experiment 1, and infants should thus rely on the last and now frequent syllable in Experiment 2.

3. Experiment 2

3.1. Materials and methods

Experiment 2 was identical to Experiment 1, except for the familiarization stream. Instead of *gi* and *vo*, which constitute the initial syllables of the object labels, the frequent syllables were *na* and *mu*, which constitute the final syllables of the object labels. The nine infrequent syllables *va*, *ga*, *pi*, *ni*, *fi*, *bo*, *vo*, *ku*, and *de* could follow the frequent syllable *mu*; and the nine infrequent syllables *da*, *ta*, *gi*, *si*, *ko*, *go*, *bu*, *pu*, and *fe* could follow the frequent syllable *na*. As for Experiment 1, the beginning and the end of the familiarization stream were ramped in amplitude.

3.1.1. Participants

Forty-one Italian 17-month old infants were tested. Seventeen were excluded for fussiness (8), equipment failure (insufficient eye tracker data, 8) or the mother not following experimental instructions (1). The remaining 24 infants (17 males, 7 females, age range: 17 month and 5 days–18 month and 7 days) were included in the final analysis.

3.2. Results

3.2.1. First look analysis

First fixations occurred on average 1067 ms after the central attractor disappeared (SD = 985 ms). Fig. 4 and Table 1 show how often infants first looked at the familiar object from the teaching phase as a function of the experimental condition. Infants did not first look significantly more often at either object in the same-infrequent (M = 1.08, SD = .78, W = 45, P = .79) or in the same frequent condition (M = .96, SD = .69, W = 30, P > .99). Infants started looking towards the familiar object from the teaching phase as often in the same-infrequent condition.
as in the same-frequent condition, Wilcoxon signed rank test, \(W = 81, P = 0.54\).

To make sure that each infant’s fixation pattern was not based on what they happened to fixate before the offset of the label, we also performed separate analyses for trials where infants already fixated an object (rather than the central attractor) before the offset of the label. In total, there were 33 such trials (34% of the trials). Among these 33 trials, fifteen occurred in the same-infrequent condition and eighteen in the same-frequent condition. After removing these 33 trials, our central pattern of results remained unchanged: infants started looking towards the familiar object from the teaching phase as often in the same-infrequent condition \((M = .75, SD = .74)\) as in the same-frequent condition \((M = .58, SD = .65)\), Wilcoxon signed rank test, \(W = 46.5, P = 0.51\).

### 3.2.2. Preferential looking analysis

Overall, the preferential looking analysis, presented in Fig. 5, confirmed the results of the first look analysis. We performed Student’s \(t\)-tests to compare the proportion of time spent looking at the familiar object in each of the five time periods of interest against chance (.5), in the same-infrequent and in the same-frequent conditions. Results showed that infants did not significantly differ from chance in either condition, for any period; \(P > .5\) after Bonferroni corrections for ten comparisons.

We ran a repeated-measures ANOVA with same-syllable (frequent or infrequent) as a within-subject factor and time as a repeated variable. One participant had no data for the first time period, so that 23 infants were included in this analysis. There was no significant effect of same-syllable, no effect of time and no significant interaction between time and same-syllable; all \(P < 1\).

Finally, planned comparisons asked whether infants looked more at the familiar object in the same-infrequent than in the same-frequent condition in each of the five time periods of interest. No significant difference was observed; all \(P > .2\).

### 3.3. Discussion

Experiment 2 differed from Experiment 1 in that the final syllables of object labels, rather than the initial syllables, occurred more frequently in the previous familiarization stream. Contrary to Experiment 1, infants did not form a stronger association with either the frequent or the infrequent syllable of a label. These results have two implications. First, infants do not systematically rely on the final syllable of a bisyllabic label, but do so only when the initial syllable is highly frequent (see also Experiment 2 in Hochmann et al., 2010).

Second, the non-referential use of frequent syllables is not independent from the syllable position. It may be that, in Italian 17-month-olds, the non-referential use is restricted to frequent syllables occurring in the initial position of a label. This may be the case if our participants acquired that knowledge from their exposure to Italian.

However, that second implication is not fully warranted yet. Indeed, both young infants (Gervain et al., 2008) and adults (Gervain et al., in preparation) tend to segment continuous streams alternating between frequent and infrequent syllables, with respect to the distribution of frequent elements in their native language. In particular, Italian-learning 7-month-olds tend to perceive an artificial speech stream, very similar to our familiarization streams, as a series of short sequences starting with frequent syllables. Thus, if we generalize the results of Gervain and colleagues to our situation, frequent elements might be perceived as initial in the familiarization of Experiment 2, whereas they appear in final position in the successive teaching and test phases. This inconsistency in position may explain the absence of effect. The frequent syllables extracted from the familiarization phase may be associated to the utterance-initial position, and infants may not recognize or be unable to use them in final position.

In Experiment 3, we asked whether the absence of a significant effect in Experiment 2 was due to a mismatch of the perceived position of frequent items in the familiarization and successive phases, or whether it implied that the effect is actually restricted to the initial position. Experiment 3 was identical to Experiment 2, but frequent-final segmentation was forced by inserting silent pauses after each frequent syllable in the familiarization. The position of frequent syllables was therefore consistently final in all phases of the experiment.

### 4. Experiment 3

#### 4.1. Materials and methods

Experiment 3 was identical to Experiment 2, except for the familiarization stream. The familiarization stream was segmented so that each frequent syllable was followed by
a phoneme-long (120 ms) silent pause. As a consequence, infants should perceive the familiarization stream as a series of sequences ending with a frequent syllable. Contrary to Experiments 1 and 2, the beginning and the end of the familiarization stream were not ramped in amplitude.

4.1.1. Participants

Forty-six Italian 17-month-old infants were tested. Twenty-two were excluded for fussiness (11), equipment failure (insufficient eye tracker data, 8), the mother not following experimental instructions (1) or side bias (2). The remaining 24 infants (17 males, 7 females, age range: 17 month and 05 days–18 month and 01 day) were included in the final analysis.

4.2. Results

4.2.1. First look analysis

First fixations occurred on average 917 ms after the central attractor disappeared ($SD = 838$ ms). Fig. 6 and Table 1 show how often infants first looked towards the familiar object from the teaching phase as a function of the experimental condition. Infants did not first look significantly more often at either object in the same-infrequent condition ($M = 1, SD = .66, W = 27.5, P = 1$). However, they were less likely than chance to look first at the familiar object in the same-frequent condition ($M = .625, SD = .65, W = 14, P = .022$). Most importantly, infants were significantly more likely to look first to the familiar object in the same-infrequent condition ($M = .625, SD = .65, W = 32, P = .039$).

To make sure that each infant’s fixation pattern was not based on what they happened to fixate before the offset of the label, we also performed separate analyses for trials where infants already fixated an object (rather than the central attractor) before the offset of the label. In total, there were 30 such trials (31% of the trials). Among these 30 trials, 13 occurred in the same-infrequent condition and 17 in the same-frequent condition. After removing these 30 trials, our central pattern of results remained unchanged: infants were significantly more likely to look first to the familiar object in the same-infrequent condition ($M = .75, SD = .68$) than in the same-frequent condition ($M = .375, SD = .49$), Wilcoxon signed rank test, $W = 12, P = 0.033$.

To compare the results of Experiments 1 and 3, we calculated for each infant the advantage for the infrequent syllable as the difference in number of first fixations to the familiar object in the same-infrequent and same-frequent conditions. The advantage for the infrequent syllable was not statistically different in Experiment 1 ($M = .43; SD = .84$) and Experiment 3 ($M = .37; SD = .82$); Wilcoxon rank sum test, $U = 558.5, P = .89$. 

Fig. 5. Results of the preferential looking analysis in Experiment 2. The x-axis represents five time periods in ms, from the disappearance of the central fixation attractor. We show the evolution of the proportion of cumulative time spent looking at the familiar object during the course of a test trial for the same-frequent condition (blue) and the same-infrequent condition (red). Error bars represent standard error of the mean.

Fig. 6. Results of the first look analysis in Experiment 3. Mean number of first looks directed towards the familiar object from the teaching phase. Error bars represent the standard error from the average. Error bars represent standard error of the mean.
4.2.2. Preferential looking analysis

The preferential looking analysis is presented in Fig. 7. We performed Student’s $t$-tests to compare the proportion of time spent looking at the familiar object in each of the five time periods of interest against chance (.5), in the same-infrequent and in the same-frequent conditions. Results showed that infants looked significantly less than chance in the same-frequent condition in the first time period; $t(23) = -3.31; P = .003$, which was significant after Bonferroni correction for ten comparisons ($\alpha = .005$). Other comparisons did not show significant differences from chance; all $Ps > .15$.

We ran a repeated-measures ANOVA with same-syllable (frequent or infrequent) as a within-subject factor and time as a repeated variable. Three participants had no data for one of the time periods in one condition, so that 21 infants were included in this analysis. The main effects of same-syllable; $F(1,20) = 1.164; P = .29$ and time; $F(4,17) = 1.61; P = .22$ were not statistically significant. There was no statistically significant interaction of these two factors; $F(4,17) = 1.94; P = .15$.

Finally, planned comparisons asked whether infants looked more at the familiar object in the same-infrequent than in the same-frequent condition in each of the five time periods of interest. The difference was significant for the first time period; $t(1,23) = 2.65; P = .014$. Other comparisons showed no statistically significant differences; all $Ps > .4$.

To compare the results of Experiments 1 and 3, we ran a repeated-measures ANOVA with same-syllable (frequent or infrequent) and time as within-subject factors and Experiment (1 or 3) as between-subject factor. Results showed a main effect of same-syllable; $F(1,40) = 6.04; P = .018$. The interaction of same-syllable and Experiment was not significant; $F(1,40) = 1.08; P = .31$. The triple interaction of same-syllable, time and Experiment was not significant; $F(4,37) = 2.16; P = .09$. No other effect or interaction were significant; all $Ps > .1$.

4.3. Discussion

In Experiment 3, contrary to Experiment 2, infants were more likely to first look at the familiar object in the same-infrequent than in the same-frequent condition. Convergent evidence was found in the preferential looking time analysis for the first time period, though not in the following periods. These results suggest that infants associated the infrequent initial syllable of a label, rather than the frequent final syllable, with a novel object.

These findings demonstrate that the absence of a frequency effect in Experiment 2 was not due to infants’ inability to infer the non-referential use of frequent items in final position. Rather, that behavior resulted from a mismatch between the perceived position of the frequent items in the familiarization and in the subsequent phases of the experiment. In Experiment 3, the familiarization stream was segmented to ensure that frequent syllables would be perceived in final position in all three phases of the experiment. Thus, not only do infants extract the frequent elements from the familiarization stream, but they also encode the position that these elements should occupy in the language (i.e., final in Experiment 3; initial in Experiments 1 and 2). The non-referential use appears to be only inferred when frequent syllables occur in the correct position.

5. General discussion

Infants in the course of language acquisition face the problem of associating chunks of speech with semantic referents. The gavagai problem, as stated by Quine (1960), exemplifies the difficulty to select the appropriate referent in the world. Here, we considered a second problem, the difficulty to identify the appropriate speech chunk to associate with a given referent. To solve this problem, we proposed that infants spontaneously infer that very frequent syllables are not referential.

In three experiments, infants were first familiarized with an artificial speech stream alternating between
frequent and infrequent syllables. Subsequently, infants were presented with an object paired with a label consisting of one frequent and one infrequent syllable. After exposure to a continuous familiarization stream, infants associated the object more strongly with the infrequent syllable than with the frequent syllable. Importantly, this happened only when the frequent syllable occurred in initial position (Experiment 1) but not when it occurred in final position (Experiment 2). If, however, infants had been familiarized with a segmented familiarization stream, so that frequent syllables were perceived in final position of bisyllabic chunks, infants associated the object more strongly with the infrequent initial syllable than with the frequent final syllable (Experiment 3).

It thus appears that infants are equipped with a mechanism leading them to treat frequent syllables as less referential than infrequent syllables. Our results further suggest that infants do not solely extract the frequent syllables from the familiarization streams, but they also encode their preferential positions (initial of final). By default, Italian infants infer that frequent syllables occupy initial positions (Gervain et al., 2008). In the face of appropriate segmentation cues, frequent syllables can however be assigned to final positions. Frequent syllables thus constitute frequent frames. This property is characteristic of function words (especially determiners), which tend to occupy a specific position relative to their lexical complement (Chan, 2008). For example, a determiner such as “the” may be represented as “theN”, where the symbol “N” can be replaced by any noun. Such frames are later used to constrain the referent of a novel word, indicating its syntactic category (Berinal et al., 2007; Brown, 1957). Combined with evidence that children and adults use frequent elements to learn positional rules (Braine, 1963, 1966; Green, 1979; Valian & Coulson, 1988; Valian & Levitt, 1996) and form distributional categories (Höhle et al., 2004; Mintz, 2006), our results further suggest that infants use frequency to recognize potential function words.

We showed that infants can infer functional properties of words from some of their distributional properties. The origin of such early linguistic knowledge should be further investigated. The 17-month-olds tested in our studies already know a number of frequent function words in Italian. One possibility is that, noticing that similar utterances such as il conigliolo (the rabbit) and un conigliolo (a rabbit) have similar referents (i.e., the concept rabbit), infants learned that a few frequent determiners (il and unin) are non-referential and subsequently generalized that property to novel frequent elements. Alternatively, the non-referential use of frequent words may be part of some core linguistic knowledge or general principle that guides language acquisition (Hochmann & Mehler, 2012).

6. Conclusion

Natural human languages are characterized by the existence of two broad classes of elements: content elements that are linked to semantic referents, and function elements that mainly serve morpho-syntax. A distributional analysis of the input allows infants to build or identify two classes of words, those that refer to entities in the world, and those that do not. Infants may then apply different learning mechanisms to each class of words, focusing on infrequent words when building the lexicon and on frequent elements when learning about syntax.

Acknowledgements

This research has been supported by McDonnell Foundation Grant No. 21002089 and the Marie Curie International Outgoing Fellowship for Career Development called “COIN – Compositionality in Infants”. We warmly thank Liuba Papeo, Mariko Moher and Jacques Mehler for comments on previous versions of the manuscript, Ansgar D. Endress for fruitful discussions on the experiments presented here, Ernő Tégla for his help in creating the stimuli and discussing the analysis, Alessio Isaia for technical support, Francesca Gandolfo and Marijana Sjkloča for administrative assistance and three anonymous reviewers for their valuable comments.

References


As briefly reviewed in the introduction, the notion of frame has been discussed in the literature concerned with word categorization. Some researchers have defined the notion of frame as consisting of two jointly occurring and non-adjacent elements (a_b; Mintz, 2003). Others have used a slightly different notion of frame, consisting only of one word or sequence of words occurring before or after the word to categorize (a_a, or _b; Braine, 1963; Fernald & Hurtado, 2006; Höhle et al., 2004). Here, we refer to the latter notion of frame.

Our experimental design, where participants need to associate labels and objects, may be tapping into the distinction between determiners and nouns, rather than the broader distinction between function and content words. Future studies should try to generalize our results to other classes of words.


