Word comprehension and production asymmetries in children and adults

Lisa Gershkoff-Stowe a,*, Erin R. Hahn b

a Department of Speech and Hearing Sciences, Indiana University, Bloomington, IN 47405, USA
b Department of Psychology, Furman University, Greenville, SC 29613, USA

ABSTRACT

Two studies investigated differences in the comprehension and production of words in 2-year-old children and adults. Study 1 compared children's speaking and understanding of the names of 12 novel objects presented over three weekly sessions. Study 2 tested adults' performance under similar training and testing conditions over two sessions. The findings indicated a comprehension advantage for both age groups. A fine-grained temporal analysis of individual words revealed that acquisition does not resemble a linear stage-wise progression from comprehension to production. Rather, dimensions of lexical knowledge develop at different rates, with words acquired, lost, and maintained over the course of learning. The findings support a dynamic and graded view of lexical processing and have implications for understanding what it means to know a word.

© 2012 Elsevier Inc. All rights reserved.

Introduction

One of the more perplexing and enduring issues in the language acquisition literature concerns the relationship between speaking and understanding. As every parent knows, young children comprehend many more words than they produce. Although such reports are merely anecdotal, they are supported by empirical studies of early vocabulary growth (Goldin-Meadow, Seligman, & Gelman, 1976; Harris, Yeeles, Chasin, & Oakley, 1995; Huttenlocher, 1974). In one widely cited study, Benedict (1979) tracked the development of eight children longitudinally from 10 to 16 months of age, using parent reports and naturalistic observation. She found that the rate of acquisition for the first 50 words in...
comprehension was nearly double the rate of acquisition in production. Children understood an average of 50 words by 13 months of age, whereas they produced only 10 words. Moreover, it took another 6 months to reach the same 50-word milestone in production. Using a much larger cross-sectional sample, Fenson and colleagues (1993) obtained normative data from children between 8 and 30 months of age. They found that even with receptive vocabularies of 150 words or more, children rarely produced the words they knew. This lexical gap continued throughout the second year, with similar asymmetries reflected in children's developing grammatical skills (Bates & Goodman, 1997).

Although our understanding of children's emerging language skills has increased considerably over the past decade or so, there has been little empirical study of the specific nature of the early word comprehension advantage and how words become coordinated in production. The goal of the current study was to provide a beginning look at changes in the ability to use words receptively versus productively. To explore this issue, we took a diachronic approach, tracking the acquisition of individual words over time. In contrast to earlier studies, our focus was on the acquisition of the words themselves rather than on the overall growth of the lexicon. This approach, although based on a restricted set of words, has the potential to yield valuable insight into the nature of the acquisition process and the mechanisms that underlie their accessibility in lexical memory for both words comprehended and words produced.

Implicit in most traditional accounts of early word learning is a view of the mental lexicon as a static entity that exists independently of its use. Words are either known or not known, entering lexical storage in an all-or-none fashion. Increasingly, however, there is evidence that the lexicon is not a "repository of facts" (Elman, 1995); rather, the nature of the information stored and retrieved is inherently dynamic and probabilistic. Several recent studies, involving both adults and children, suggest that the learning and retention of words is better characterized as a graded process (Smith & Yu, 2008; Vouloumanos, 2008). For instance, Smith and Yu (2008) showed that lexical knowledge is built incrementally from the co-occurrence statistics between words and objects across multiple instances. Accordingly, it should be possible to capture partial states of knowledge during the course of children's word acquisition.

In the current study, we adopted a simple explanatory framework of word learning in which information is processed via the spread of activation through an organized and richly connected network of individual units (McClelland & Rumelhart, 1981). Lexical processing is assumed to consist of two levels of interconnecting units corresponding to the sound form and meaning of words. Activation flows bidirectionally between levels as well as laterally within each level. These interconnections can be both facilitative and inhibitory. Word learning is conceptualized as occurring in two ways: (a) through the accrual of activation each time a word is encountered and (b) through system-wide changes in the organization of words as the size or density of the lexicon increases (Gershkoff-Stowe & Hahn, 2007). On this view, graded activation operates incrementally and probabilistically on both the lexical representations themselves and the processes that underlie their retrieval. A key premise of this interactive system is that the stronger the activation of a word, the greater its accessibility.

Here we investigated three main questions. First, how does repeated experience with a word affect its strength of activation, and hence accessibility, in comprehension versus production? We addressed this question by presenting children with novel names for novel objects over three consecutive weekly training sessions. At each session, children's knowledge of the experimental words was tested in both a recognition task and a naming task. This microgenetic approach provides a means for studying incremental changes from session to session and allows the examination of individual differences in word learning and retrieval skills (Siegler, 2006). In addition, this approach allows the possibility of discerning changes in the relationship between receptive and productive word knowledge. A number of different learning patterns have been described in the literature, with the most common being the acquisition of a word first in comprehension and then in production (Clark, 1993). However, it is also possible for production to precede comprehension or for both to emerge simultaneously. In the current study, we looked at the frequency of the asymmetries between production and comprehension and whether the relationship changes over time.

A second question we asked is how productive vocabulary size influences the comprehension–production gap. Several studies have noted that once children achieve a critical mass of words in their spoken lexicon—typically between 50 and 100 words—there is a striking change in the speed with
which they learn new words (Bloom, 1973; Gershkoff-Stowe & Smith, 1997; Marchman & Bates, 1994; see also network simulations of vocabulary growth by Plunkett, Sinha, Moller, & Strandsby, 1992). Whereas words were once learned slowly and on the basis of many repetitions, words now are acquired rapidly, often after a single exposure. Other research has shown that concurrent to this initial period of accelerated growth in expressive vocabulary, toddlers also make dramatic advances in receptive language. For instance, Fernald, Perfors, and Marchman (2006) found significant gains in speech processing efficiency from 15 to 25 months of age, with older infants responding faster and more reliably on word recognition tasks than younger infants. In studies of Spanish–English bilinguals, Marchman, Fernald, and Hurtado (2010) found that the speed with which children identified the referents of familiar words was linked to the overall number of words they produced in both languages. Taken together, the evidence suggests that language understanding and language production grow in tandem and come to support increasingly complex lexical systems.

Accordingly, we predicted that children with many words in their productive lexicon would show more robust learning in both comprehension and production than children with fewer words. Learning that is robust would be signaled by faster rates of acquisition, more rapid coordination between comprehension and production, and greater stability of learning over time. Naturalistic observation suggests that word learning is often accompanied by a high degree of instability or flux. For example, Bloom (1974) noted that certain words children had produced earlier in her longitudinal study could not be reliably elicited on subsequent occasions.

Meara (2009) developed an innovative method for predicting long-term vocabulary changes in adult second language learners using a transitional probability matrix. In his model, vocabulary moves fluidly between discrete knowledge states. Learning is represented at two separate points in time, revealing the nature of the change between them. Adapting the model to the current design, we studied individual children’s stability of words by looking at movement from Session 1 to Session 2 and from Session 2 to Session 3 across four states of knowledge: no knowledge of the word, knowledge in comprehension alone, knowledge in production alone, and knowledge in both comprehension and production. These values can be plotted in two 4 x 4 transition matrices for each child, and the proportion of words in any one state of knowledge can be calculated from session to session. Thus, it is possible to determine how much growth, attrition, and stability children exhibit in their word knowledge over time.

A final question we addressed in the current research is whether the predicted comprehension advantage in children is driven by the cognitive or physical immaturity of a nascent system (e.g., limitations in working memory capacity, articulatory control). By this account, age alone induces qualitative changes in the processes that underlie word learning and retrieval (for a discussion on theories linking maturational constraints to changes in vocabulary acquisition, see Snedeker, Geren, & Shafto, 2012). Alternatively, it may be that lexical processes are essentially similar across development and that the comprehension–production asymmetry reflects differences in the way information is accessed in lexical memory.

With respect to the latter hypothesis, it is generally accepted that speaking and understanding depend on similar processing mechanisms but that the flow of information differs (e.g., Cutting, 2009). In comprehension, the listener must interpret the acoustic pattern of input and map the sound form onto existing meaning. Production, on the other hand, begins with a semantic representation, which then maps to the corresponding phonological form. Differences in the starting points of comprehension and production implicate differences in retrieval demands. Huttenlocher (1974) proposed that even shallow and incomplete storage of an existing word form might be “good enough” to permit access to its meaning. In contrast, word production requires the retrieval of phonological information that might not be needed in similar levels of detail in comprehension (Bock & Griffin, 2000). These disparities may affect how lexical information is selected and retrieved.

If age-related abilities underlie the lexical gap, then we should find that children, but not adults, perform better on measures of comprehension relative to production when tested on their knowledge of novel words. Alternatively, if even highly proficient learners acquire new words in comprehension and production at different rates, then this would suggest that the disparity stems not from maturational factors alone but rather from differences in the lexical processing mechanisms that support the two systems. Accordingly, we compared the language performances of children and adults under
similar word learning conditions. Two studies were conducted. Study 1 examined the comprehension–production gap in 2-year-old children over three consecutive weekly sessions. Study 2 tested adult word learners in a similar training procedure at two sessions a week apart.

Study 1

In Study 1, we presented toddlers with novel object labels and tested their ability to recognize and name the words at weekly intervals. The purpose was to track the expected asymmetry between words acquired in comprehension and production and to investigate how productive vocabulary size might mediate the gap. Three aspects of the acquisition process were measured, providing a detailed examination of the comprehension–production relationship. First, we looked at the number and order of acquisition of individual words in the recognition and naming tasks across three training sessions. Second, we calculated the lag time for those words that were successfully acquired in both comprehension and production. Third, we studied the trajectory of individual words using a developmental matrix that captured the complex and dynamic nature of the children’s word acquisitions in terms of the numbers of words lost, gained, and remaining stable from week to week.

Method

Participants

The participants were 12 24-month-olds (8 boys and 4 girls). The children were typically developing and from middle-class families in which English was the only language spoken. Ages of the children at the start of the study ranged from 23 months 9 days to 24 months 22 days, with a mean of 24.11 months. Children’s spoken language vocabulary was assessed a week prior to the start of the study using the Words and Grammar Scale of the MacArthur–Bates Communicative Development Inventories (Fenson et al., 1993). As expected, there was large variation among individual children, with parent reports ranging from a low of 37 words to a high of 523 words ($M = 280$ words).

Stimuli

The stimuli were 12 novel objects created for the purposes of the experiment. Sample objects are shown in Fig. 1. The objects varied in size ($M = 10.9$ cm) and material; all were easy for young children to grasp and manipulate. Colored pictures of the objects (~15 × 10 cm) were used in the recognition task. Pictures were attached to white poster board (~38 × 76 cm). Novel words were created from the phonetic inventory that children typically use when beginning to produce words (Vihman & Miller, 1988). All words consisted of phonotactically legal strings that were either one or two syllables in length (e.g., pib, booma).

Fig. 1. Sample stimuli.
Procedure

Individual children and their parents visited the laboratory once a week for three successive weeks. At each session, children were presented with the same set of 12 experimental objects, one at a time, in a unique predetermined random order. During training, parents were encouraged to draw their children’s attention to the objects by pointing to and repeating the labels initially named by the experimenter. Naming occurred in an ostensive context when both children and adults were jointly focused on the referent (e.g., “Look! Here’s the booma. See the booma? This is a booma! Show mommy the booma.”). Children were encouraged to manipulate the objects and to repeat the labels provided by the adults. Together, the experimenter and parents named each of the objects 10 times on average ($SD = 2.87$).

Immediately following the training phase, children were tested on their comprehension and production of the object words. Recognition consisted of a forced-choice selection task. Children were presented with pictures of 6 of the objects at a time and were asked to point to each one the experimenter named (e.g., “Where’s the booma?”). The pictures were randomly displayed on a small poster board placed directly in front of the children. The same procedure was repeated with the remaining 6 objects. In the naming task, the experimenter held up one three-dimensional object at a time and asked children to say the name (“What’s this?”). The order of individual items was randomly presented within each task, and corrective feedback was provided. Thus, children heard the experimenter label each object an additional two or three times during testing at each session. Comprehension testing always preceded testing in production. This order of testing provided children with additional instances of naming in the comprehension test prior to production testing. By conferring a possible advantage on children’s performance in the naming task, the design presents a more rigorous test of the hypothesis that word comprehension precedes word production.

Coding

All sessions were recorded for subsequent coding. Comprehension was credited if children pointed to the correct picture of the object. Only initial responses in the recognition task were counted. Production was credited if children said the correct word or approximated the target. An approximated utterance was one that contained at least two of the target word’s phonemes in the correct position (e.g., oop for noop). Reliability was established for 25% of the children randomly selected at each session. Agreement between two coders was 91.6% (range = 83–100) for comprehension and 93.6% (range = 92–100) for production. Comprehension scores were lower than expected because children sometimes generated rapid pointing strings to multiple pictures during a single trial. Differences were resolved by reviewing the recordings and through discussion.

Results

The lexical gap

We first document the predicted lexical asymmetry for all 12 children across the three weekly sessions. Children’s knowledge of the 12 experimental words is presented in Fig. 2 as the mean number of correct responses. As expected, word learning improved in both the recognition task and naming task with additional exposure at each training session. This was true for every child, although advances occurred at different rates. On average, children correctly recognized 3.25 words at Session 1 and 6.08 words at Session 3. In contrast, they rarely generated labels, even with extended practice. Children produced an average of less than 1 word ($M = 0.42$) at Session 1 and only 2.67 words by Session 3.

A $2 \times 3$ repeated measures analysis of variance (ANOVA) was conducted, with task type (comprehension or production) and session as within-participant variables. This analysis yielded the predicted main effect of task type, $F(1,11) = 137.84, p < .001, \eta^2_p = .93$ (children learned significantly more words in comprehension [$M = 4.75$] than in production [$M = 1.61$]) and a main effect of session, $F(2,22) = 14.53, p < .001, \eta^2_p = .57$. Tukey’s HSD (honestly significant difference) post hoc analyses revealed reliable increases at all sessions of testing ($p < .05$). There was no interaction between task type and session.
Additional analysis examining the relationship between the number of times the experimenter and parents named the words across the three sessions and children's learning of the words revealed no significant correlation for either comprehension \( (r = -.12, p = .71) \) or production \( (r = -.20, p = .53) \).

**Temporal relationship between comprehension and production**

Even after three sessions of repeated training, nearly half of the words children correctly identified in comprehension failed to be produced \( (43.75\%) \). In contrast, children rarely said a word that they did not also understand \( (4.17\%) \); rather, children generally comprehended those words that they also produced \( (26.39\%) \). To determine how individual words might be sequentially linked in comprehension and production, we examined the order of acquisition for the words that children had successfully recognized and named by Session 3, a total of 38 words representing 11 children \( (1 \text{ child failed to produce any of the experimental words}) \). Three patterns of order acquisition were possible, namely that comprehension of a word might precede production \( (C \rightarrow P, n = 20) \), production might precede comprehension \( (P \rightarrow C, n = 5) \), or comprehension and production might occur within the same session \( (C&P, n = 13) \).

A chi-square goodness-of-fit test was conducted to compare differences across the three orders of acquisition. This analysis yielded a significant effect, \( \chi^2(2, N = 38) = 8.89, p < .05, \Phi = 1.44 \). We used standardized residuals to identify which patterns of acquisition differed significantly from chance. We found that the frequency of \( C \rightarrow P \) words was significantly greater than expected by chance \( (\text{standardized residual} = 2.11, p < .05) \), and the frequency of \( P \rightarrow C \) words was significantly less than expected by chance \( (\text{standardized residual} = -2.13, p < .05) \). Thus, the data support what has been generally assumed, namely that most words that children acquire appear to be sequentially linked, reaching threshold first in comprehension and then in production. However, we also found that a sizable minority of words do not exhibit a comprehension advantage.

**Word learning in children with large and small productive vocabularies**

How might vocabulary size affect the progression from comprehension to production? We hypothesized that the experimental words would be learned more readily among children who were reported to have relatively large productive vocabularies, with the idea being that as more words become known to children, lexical processing efficiency increases. Based on parental report, we used a median split to create two groups: high and low productive vocabularies (median = 360 words). The 6 children in the High Vocabulary group had a mean score on the MacArthur–Bates Communicative Development Inventories \( (\text{Fenson et al., 1993}) \) of 405 words \( (SD = 80.6) \). The remaining 6 children in the Low Vocabulary group had a mean score of 155 words \( (SD = 113.9) \). An independent \( t \) test confirmed that the difference was statistically reliable, \( t(10) = 4.38, p = .001, d = 2.53 \).
To test our prediction concerning the effect of vocabulary size on word learning, a $2 \times 2 \times 3$ mixed ANOVA with task type and session as within-participant variables and vocabulary as a between-participant variable was conducted on the frequency of children's correct responses. The three-way interaction was significant, $F(2,20) = 7.44, p = .004, \eta^2_p = .427,$ as were the main effects of session and task type. To analyze the significant interaction, we used simple effects. First, we divided the data by task type so that we could examine the Vocabulary \times Session interaction for recognition and naming separately. With respect to performance on the comprehension task, the two-way interaction reached significance, $F(2,20) = 4.03, p = .034, \eta^2_p = .120$ (see Fig. 3A). We then compared the High Vocabulary and Low Vocabulary groups at each of the three sessions. The tests revealed that children in the High Vocabulary group understood more of the experimental labels than children in the Low Vocabulary group at Sessions 1 and 2. At Session 3, however, comprehension scores were equivalent between the two groups.

When the same analysis was repeated by isolating performance on the production task, the two-way Vocabulary \times Session interaction approached significance, $F(2,20) = 3.47, p = .051, \eta^2_p = .146$ (see Fig. 3B). We found no difference between the numbers of words children produced at Session 1. In contrast, at Sessions 2 and 3, children in the High Vocabulary group said more of the experimental words relative to children in the Low Vocabulary group.

Our second measure of word learning efficiency involved the temporal properties of the comprehension–production gap determined by the lag (quantified by the number of sessions) between words acquired in both comprehension and production. We hypothesized that children in the Low Vocabulary group would benefit more from extra time to strengthen production processes than children in the High Vocabulary group. Thus, we expected children in the Low Vocabulary group to exhibit a wider lexical gap initially.

**Fig. 3.** (A) Mean numbers of words correct in comprehension by session for children in the High Vocabulary and Low Vocabulary groups. (B) Mean numbers of words correct in production by session for children in the High Vocabulary and Low Vocabulary groups.
To calculate the time lag between words comprehended and words produced, we aligned individual children’s data by the session of first production and examined comprehension of the word at each prior session. The data, presented in Table 1, consisted of 27 words for children in the High Vocabulary group and 11 words for children in the Low Vocabulary group. Positive lags (Lag +1 and Lag +2) refer to words that children comprehended prior to production; negative lags (Lag −1 and Lag −2) indicate words that children produced prior to comprehension. The No Lag category refers to words that were comprehended and produced simultaneously. To take into account differences in base rates, Table 1 includes the total number of possible words that could have been classified in each lag category. For example, a word initially produced at Session 3 could have had a lag of zero, +1, or +2, whereas a word initially produced at Session 1 could only have had a lag of zero. We then calculated proportions using the base rates.

Although the analysis yielded too few spontaneous productions to support inferential statistics, the data suggest that productive vocabulary size may mediate differences in the temporal relationship between the comprehension and production of individual words. Contrary to expectations, children in the Low Vocabulary group were more likely to produce a word soon after hearing it, whereas children in the High Vocabulary group took longer to coordinate production of a word. Thus, children in the High Vocabulary group appeared to experience an extended “silent period” (Gibbons, 1985) during which they encoded information about the meaning of a word but failed to use the word themselves. Inconsistent with this result, however, was the additional finding that 4 of the 6 children in the High Vocabulary group had at least one instance of producing a word prior to comprehension; no children in the Low Vocabulary group exhibited this unusual pattern.

**Stability of words in comprehension and production**

A final measure of word learning efficiency concerned the reliability of words retained in lexical memory. We found that children’s knowledge of individual words often fluctuated from week to week, with some words appearing and then disappearing—at times in comprehension, at times in production, or at times in both. Accordingly, we attempted to capture the dynamic nature of the acquisition process by tracking each word across the three sessions of training for each child. Our initial prediction had been that children’s productive vocabulary size would matter, with higher vocabularies showing greater stability than lower vocabularies.

We created developmental matrices (Meara, 2009) to capture the movement of individual words in and out of particular states of knowledge. Because testing occurred at each of three sessions, two developmental matrices (each representing the transition from one consecutive session to the next) were constructed for each child. Fig. 4 presents the data for individual children separately by vocabulary group. Each horizontal row within a matrix indicates the number of words in a particular knowledge state at the earlier session. To illustrate, Child 1 showed no knowledge (Ø) of a total of 10 words at Session 1, whereas 1 word was learned in comprehension only (C) and an additional word was learned in both comprehension and production (C + P), indicated in the second and fourth rows,

| Table 1 | Word comprehension–production lag in children as a function of vocabulary size. |
|---------|-------------------------------|-------------------|------------------|
|         | Possible words | Words | Proportion |
| High Vocabulary | | | |
| Lag +2 | 12 | 6 | .50 |
| Lag +1 | 21 | 9 | .43 |
| No Lag | 27 | 7 | .26 |
| Lag −1 | 5 | 4 | .80 |
| Lag −2 | 3 | 1 | .33 |
| Low Vocabulary | | | |
| Lag +2 | 6 | 0 | .00 |
| Lag +1 | 10 | 5 | .50 |
| No Lag | 11 | 6 | .55 |
| Lag −1 | 0 | 0 | – |
| Lag −2 | 0 | 0 | – |
respectively. Note that the numbers total to 12, which is the total number of words presented per session. Looking next at the vertical columns within the matrix, the number of words in a given knowledge state at the following session are also indicated. For example, we can see that Child 1 at Session 2 showed no evidence of knowing (Ø) a total of 4 words, whereas an additional 5 words were learned in
comprehension only (C) and 3 words were learned in both comprehension and production (C + P). Again, these numbers account for all 12 words presented at Session 2.

Importantly, the cells at the intersection of a row and column provide information about the changing knowledge states of individual words. The unshaded diagonal contains the number of words that exhibited stability from one session to the next, including whether the word was known to the child or unknown. For example, the first matrix of Child 1 in Fig. 4 shows that a total of 6 words remained stable between Session 1 and Session 2: 1 in comprehension, 1 in comprehension and production, and the remaining 4 as unknowns. The light gray cells above the diagonal contain the words that were gained from one session to the next, whereas the dark gray cells below the diagonal contain the words that were lost from one session to the next. As shown, an additional 6 words were gained at Session 2, with 4 words that were previously unknown being comprehended and 2 words that were unknown in Session 1 being recognized and produced in Session 2. A similar analysis occurred from Session 2 to Session 3. Looking at Fig. 4 as a whole, what is particularly striking are the different developmental patterns across the 12 children independent of vocabulary size.

From these matrices, we calculated the proportions of words that were gained, were lost, and remained stable at each of the two time periods. Table 2 summarizes the data, showing that the learning trajectories for individual children are distinctly unique. In addition, independent t tests revealed no significant differences between children in the Low Vocabulary and High Vocabulary groups, although the attrition rate of words at Transition 1 was marginally significant, \( t(10) = 1.89, p = .087 \). Specifically, children in the High Vocabulary group (\( M = .18 \)) lost more words from Session 1 to Session 2 than children in the Low Vocabulary group (\( M = .07 \)).

### Discussion

The conventional view of vocabulary growth is one of steady progression from initial understanding of a word to full productive use. Our goal in Study 1 was to obtain a fine-grained temporal analysis of the learning trajectory for 2-year-olds trained and tested on a set of 12 new object words over a 3-week period. The results clearly demonstrated the anticipated word comprehension–production gap; all children, regardless of vocabulary size, understood more experimental words than they produced at each of the three sessions. However, we also found diverse patterns of learning and attrition. Gains were not steady; children experienced temporary stops and spurts, sometimes understanding or

<table>
<thead>
<tr>
<th>Group</th>
<th>Word stability</th>
<th>Transition 1</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Stability</td>
<td>Growth</td>
<td>Attrition</td>
<td>Stability</td>
<td>Growth</td>
<td>Attrition</td>
<td></td>
</tr>
<tr>
<td>Low Vocabulary</td>
<td></td>
<td>.50</td>
<td>.50</td>
<td>.00</td>
<td>.33</td>
<td>.25</td>
<td>.42</td>
<td></td>
</tr>
<tr>
<td>Child 1</td>
<td>.50</td>
<td>.33</td>
<td>.00</td>
<td>.42</td>
<td>.42</td>
<td>.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child 3</td>
<td>.92</td>
<td>.08</td>
<td>.00</td>
<td>.92</td>
<td>.08</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child 4</td>
<td>.83</td>
<td>.33</td>
<td>.33</td>
<td>.42</td>
<td>.58</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child 5</td>
<td>.58</td>
<td>.33</td>
<td>.33</td>
<td>.42</td>
<td>.58</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child 6</td>
<td>.58</td>
<td>.29</td>
<td>.07</td>
<td>.30</td>
<td>.33</td>
<td>.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>.64</td>
<td>.25</td>
<td>.17</td>
<td>.17</td>
<td>.33</td>
<td>.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High Vocabulary</td>
<td></td>
<td>.58</td>
<td>.25</td>
<td>.17</td>
<td>.50</td>
<td>.33</td>
<td>.17</td>
<td></td>
</tr>
<tr>
<td>Child 11</td>
<td>.67</td>
<td>.25</td>
<td>.08</td>
<td>.33</td>
<td>.50</td>
<td>.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child 12</td>
<td>.50</td>
<td>.33</td>
<td>.17</td>
<td>.58</td>
<td>.33</td>
<td>.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>.49</td>
<td>.33</td>
<td>.18</td>
<td>.49</td>
<td>.32</td>
<td>.20</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
saying words that later dropped out of use. Such behavior is characteristic of regression, or U-shaped development, a phenomenon that tends to mark periods of transition (Gershkoff-Stowe & Thelen, 2004). These facts together suggest that 2-year-old word learners do not progress through a series of well-defined predictable stages of word acquisition.

We also found that the size of children’s productive vocabularies, as measured by the MacArthur–Bates Communicative Development Inventories, was associated with differential rates of word learning in both comprehension and production. Children with large expressive vocabularies tended to be rapid comprehenders and more precocious overall relative to children with smaller expressive vocabularies. Qualitative differences in the pattern of lag were also observed. Surprisingly, however, the direction of this difference was contrary to predictions. Children with many words in their productive vocabularies took more time to say the words they knew in comprehension compared with children with fewer words.

One possible explanation for the observed pattern of findings is that children with sizable lexicons require less activation to recognize the correct referent of a word, but require more activation to generate a word for naming, relative to children with smaller vocabularies. This may be because of increases in semantic neighborhood density as more words are acquired at a rapid rate (Gershkoff-Stowe & Hahn, 2007). In particular, as suggested by studies of adult lexical processing, semantically related neighbors support access in comprehension but may function as competitors and impede lexical access in production (Neely, 1991).

It is also the case that differences in the requirements of the task are likely to affect the ease with which information is processed. In comprehension, two kinds of inputs jointly activate lexical knowledge: the object and the word. In production, only one input—the object—is available. This suggests that multiple retrieval cues present at the time of comprehension testing may help to focus attention on the key aspects of the object to be identified, leaving more resources available for encoding the target word. As such, individual differences in attentional resources between children in the High Vocabulary and Low Vocabulary groups may also contribute to unique patterns of learning.

In sum, the current findings reveal that in young children, newly acquired words are fragile and subject to considerable flux. In Study 2, we investigated whether adult word learners display a more stable pattern of acquisition from one testing session to the next and whether they also show a characteristic comprehension–production gap. If they do, we can conclude that the comprehension advantage is not solely attributable to the maturational status of young word learners. Rather, the source of the disparity is likely to reside in the basic processing mechanisms that underlie the retrieval of words in comprehension versus production.

### Study 2

**Method**

**Participants**

A total of 20 undergraduates (15 women and 5 men) from Furman University were recruited to participate. All were native speakers of English and ranged from 18 to 22 years of age.

**Stimuli and procedure**

Adults were trained and tested individually using the same set of 12 experimental words as in Study 1. The procedure was similar but was modified in several ways to take into consideration differences in linguistic and general knowledge of the world. First, participants attended only two weekly sessions in anticipation of possible ceiling effects on word learning performance. Second, adults saw all 12 pictures at one time during comprehension testing to minimize the opportunity to recruit higher level cognitive strategies. Third, we implemented a more conservative coding procedure such that the correct production of a word required an exact phonetic match.

The final procedural difference concerned the order of testing. In Study 1, we maximized the likelihood that children would produce the words by testing first in comprehension and then in production. This decision was based on previous studies of 2-year-olds that commonly yield floor effects.
when children are asked to produce new labels prior to recognition testing (Childers & Tomasello, 2002; Hahn & Gershkoff-Stowe, 2010). In the case of adults, however, we anticipated less need to augment productive word learning. As a result, we counterbalanced the testing order; half of the adults were tested first in comprehension (as in Study 1), whereas the remaining participants were tested first in production. This additional manipulation provided an opportunity to test whether the nature of the relationship between comprehension and production was robust enough to withstand minor variations in exposure.

A random group of 10 participants was selected to establish reliability across the two sessions. Interrater agreement between two coders was 100% for comprehension and 97% for production.

Results

The lexical gap

The principal question we addressed in this analysis is whether adults learning a novel set of words exhibit a comprehension–production asymmetry comparable to that of 2-year-olds. Fig. 5 presents the mean numbers of correct responses in comprehension and production over the two sessions. Similar to the children trained and tested in Study 1, the adults in this study improved in both tasks with additional training. Also like the children, the adults showed a word comprehension advantage, most notably at Session 1.

We expected that testing order would affect learning such that additional exposure during the first test would improve performance on the second test. Accordingly, we submitted the data to a $2 \times 2 \times 2$ mixed ANOVA with the between-participant factor of order (comprehension testing first or production testing first) and within-participant factors of task type (comprehension or production) and session. The three-way interaction was significant, $F(1,17) = 5.71, p = .029, \eta^2_p = .251$. We decomposed the interaction by separating the data according to session and then using simple effects to analyze the resulting Task Type × Order interactions.

Looking first at performance at Session 1 (see Fig. 6A), the Task Type × Order interaction reached significance, $F(1,17) = 4.51, p < .001, \eta^2_p = .539$. Subsequent analysis using simple effects indicated that, as predicted, performance on the second test was enhanced by the additional exposure that occurred with the first test. That is, when production was elicited before recognition, comprehension scores were higher; when comprehension was tested first, production scores were higher. However, order mattered only at Session 1. When the Session 2 data were examined, neither the Task Type × Order interaction nor the main effect of order reached significance ($ps > .10$) (see Fig. 6B). Only the main effect of task type was reliable, $F(1,17) = 21.37, p < .001, \eta^2_p = .117$, indicating that comprehension ($M = 11.26$) continued to outstrip production ($M = 9.51$) at Session 2.

Together, the findings indicate that, like the 2-year-olds in Study 1, adult word learners experienced a comprehension–production gap during the course of learning a small set of novel words. Moreover, the gap occurred regardless of whether adults were exposed to extra practice through...
Temporal relationship between comprehension and production

By the end of the second session of training, adults had recognized and/or named all but 4.2% of the experimental words. Further breakdown revealed that the majority of the words learned were both comprehended and produced (75.8%). In contrast, few words were learned in comprehension alone (17.5%), and even fewer words were learned in production alone (2.5%). As in Study 1, we analyzed how the words that were jointly comprehended and produced \((N = 186, \text{range} = 5–12)\) were sequentially linked across the two sessions. We analyzed the same three patterns of order acquisition that we used in Study 1: comprehension before production (C \(\rightarrow\) P), comprehension and production within the same session (C&P), and production before comprehension (P \(\rightarrow\) C).

We conducted a chi-square test of independence to compare differences across the three types of acquisition patterns as a function of testing order (production first vs. comprehension first). The analysis revealed a significant effect, \(\chi^2(2, N = 186) = 23.17, p < .01, \Phi = .35\). Specifically, adults who were tested first in production acquired more C \(\rightarrow\) P words (standardized residual = +2.82, \(p < .01\)). In contrast, when production was tested after comprehension, participants acquired fewer C \(\rightarrow\) P words (standardized residual = −2.70, \(p < .01\)). No other standardized residuals reached significance.

Stability of words in comprehension and production

We also examined the trajectory of individual words from Session 1 to Session 2 using separate matrices for each testing order (Meara, 2009). Unlike the child data, however, Fig. 7A and B present the data averaged across the 10 adult participants in each condition. We chose to compress the
information provided by the individual matrices in light of the fact that adults within a condition showed little between-participant variability. Nor was there much within-participant fluctuation, particularly relative to the children in Study 1. Rather, interesting differences emerged between the two testing orders, as suggested by the previous analyses.

Given that testing occurred at two sessions, only one matrix for each condition is needed to illustrate changes in the learners’ state of knowledge. As before, the unshaded diagonal within the matrix contains the number of words that remained stable from one session to the next. The light gray cells above the diagonal contain the words that were gained from one session to the next, whereas the dark gray cells below the diagonal contain the words that were lost from one session to the next.

Adults in Fig. 7A were tested first in comprehension and then in production, whereas those in Fig. 7B had the reverse testing order. This latter group presents a different profile that reflects the relative strengthening of words in comprehension as a result of additional exposure during production testing. Although there was no difference in the overall proportions of words gained, retained, and lost between the two groups, as shown in Table 3, the route for acquiring words jointly in comprehension and production differed. This is seen by looking at the first two rows of the C + P column above the diagonal in gray. As shown, adults in the comprehension before production condition were equally likely to learn a word jointly in comprehension and production (C + P) at Session 2 either by moving from not knowing the word (Ø) (M = 2.0) at Session 1 or by moving from knowing the word in comprehension only (C) (M = 2.0). In contrast, adults in the production before comprehension condition were significantly more likely to learn a word jointly in comprehension and production by moving from knowing a word in comprehension only (M = 4.3) than by moving from not knowing the word (M = .30), t(9) = 8.48, p < .001.

Also striking is the fact that adults rarely lost a word in either test order condition. Rather, words learned previously in both comprehension and production exhibited considerable stability. In this study, 5 adults accounted for a total of 7 words (2.9%) lost from Session 1 to Session 2, with 4 of those words being lost in comprehension. By comparison, for the 2-year-olds in Study 1, most of the changes from Session 1 to Session 2 involved adding new words in comprehension while simultaneously losing other words that had been previously recognized but not yet produced.

![Fig. 7](image_url)

**Fig. 7.** (A) Mean stability matrix for adults in the comprehension before production test condition. (B) Mean stability matrix for adults in the production before comprehension test condition.
Discussion

The central finding of Study 2 is that adults demonstrated a comprehension–production gap during the course of learning a small set of novel words. Like the toddlers in Study 1, knowledge of a word preceded the ability to use it. We interpret this as evidence that the underlying operations of comprehension and production are fundamentally similar in adults and children. The lexical gap is not merely a consequence of developmental changes in cognitive or oral motor functioning; rather, differences in the number of words recognized and named reflect differences in the processing mechanisms involved in accessing words for speaking and understanding. Although the findings underscore important commonalities between adults and toddlers, we also found substantial age-related differences in the speed and stability of learning.

A second important finding in Study 2 is that the word gap between comprehension and production remained irrespective of the order in which recognition and naming were tested. As expected, the additional exposure that adult participants received from the first test within a session supported subsequent learning on the second test. Specifically, participants who were tested initially in recognition were more likely to comprehend and produce words—particularly at Session 1—than participants who were tested initially in production. By Session 2, however, both groups were comparable in terms of the total number of words correctly recognized and produced.

General discussion

Why does knowledge of a word sometimes appear to be independent of a speaker’s ability to use it? The purpose of the current research was to address this question by focusing on changes in the accessibility of words learned in comprehension and production in 2-year-old children and adults. The results showed clear evidence of learning. However, growth in receptive vocabulary was not equivalent to growth in productive vocabulary. Adults as well as children needed more time to learn
novel words for speaking than for understanding; this gap subsequently narrowed for both age groups with additional training. Although the lag between words comprehended and words produced in children is predicted by other findings in the literature (Benedict, 1979; Clark & Hecht, 1983), little is known about the growth patterns of individual words as children are learning them. Moreover, the discovery that adults exhibit a similar but less pronounced lag calls for further examination of this long-standing question.

We used a microgenetic design to explore the nature of the relationship between lexical comprehension and production. This approach allowed us to investigate systematically the effects of repeated encounters with a word. The findings indicated a wide range of learning patterns, with no consistent hierarchical connection between words learned initially in comprehension and then in production. This fact suggests that advances in word production skills do not depend on a certain level of skill in comprehension, nor do individual words need to stabilize in comprehension before speakers are able to produce them.

The use of a developmental matrix to examine individual differences provided a valuable tool for investigating the nonlinear progression of words through various states of partial knowledge. Two learners may show the same increase in growth from the beginning to the final session of testing but may exhibit unique developmental trajectories. Analyzing individual children using the matrix enabled us to detect incremental changes that are otherwise obscured in studies using mean scores to make broad generalizations about groups. We found that newly acquired words often slipped in and out of a child’s lexical repertoire even after the child demonstrated mastery in comprehension and production. This result stands in contrast to adult word learners who exhibited considerable stability in their lexical knowledge. Given the qualitative nature of the stability analysis, however, these results should be interpreted with caution.

Thelen and Smith (1994) suggested that behavior variability is an important component in the process of development and serves as a marker of changes in underlying processes (see also van Geert & van Dijk, 2002). In the current study, the variability observed in young children’s knowledge of words in comprehension and production likely reflects the complexity of retrieval processes in a rapidly changing lexicon. As such, the results reveal an early fragility of word retrieval processes and lend support to a graded view of lexical development (McMurray, Samelson, Lee, & Tomblin, 2010). By this view, acquisition of a word is not an all-or-none event. Rather, the strength of a word depends on its surrounding relationship to other similar words; on here-and-now contextual influences (e.g., recently hearing or saying a word, momentary lapses of attention), and on the child’s current level of cognitive functioning, that is, operations such as perception and memory that support the acquisition and maintenance of word knowledge (Munakata, 2001).

{\textit{Lexical access in comprehension and production}}

Any general theory of vocabulary acquisition must account for the intra-individual variability and nonlinear patterns of learning observed in the current study. The results strongly suggest that 2-year-olds learn and retrieve words in comprehension and production in stochastic fashion, which is realized as relative strengths of activation. Although the current study was not designed to isolate the factors affecting the rate and order of acquisition, a number of other studies have shown that retrieval success is subject to several probabilistic factors such as word frequency, neighborhood density, and phonotactic probability. Together, these factors affect how activation is propagated through the system, as indexed by the rate at which words accumulate activation, the strength of competition effects, and the ability to resist interference (Dell, Schwartz, Martin, Saffran, & Gagnon, 1997).

Importantly, many of these lexical factors differentially influence the speed and accuracy of words in recognition and naming tasks, thereby underscoring differences in the corresponding operations that underlie them. For example, adults are more accurate and faster to produce low-frequency words from high-density phonological neighborhoods than low-frequency words from low-density neighborhoods (Vitevitch & Sommers, 2003). This effect is reversed, however, in tasks of word recognition. Storkel and colleagues (Storkel, 2009; Storkel, Armbruster, & Hogan, 2006) also found that both adults and children are more likely to learn words that come from dense phonological neighborhoods than from sparse neighborhoods and to learn rare sound sequences more readily than common sound
sequences. Still other studies reveal that the richness and connectivity of semantic representations play a significant role in facilitating word learning and retrieval (Capone & McGregor, 2005). Particularly among young children, the strength and number of pathways among semantically related words is rapidly changing over time and, thus, is likely to have consequences for how words are stored and retrieved in the developing lexicon.

We expected vocabulary size to be an important variable in influencing successful recognition and naming. Accordingly, Study 1 examined children's lexical retrieval skills in light of their underlying vocabulary development. Although we found that children with more words in their productive lexicon acquired words in comprehension at a faster rate than children with fewer words, there was no difference between the two groups by Session 3. Furthermore, children in the High Vocabulary group generally took longer to coordinate production of the words they knew in comprehension than children in the Low Vocabulary group.

The developmental patterns observed among children in Study 1 are consistent with a dynamic view of lexical acquisition in which successful retrieval emerges from activation processes that operate in the moment-to-moment activity of seeing, hearing, and naming objects. As more learning occurs, individual units are strengthened, which in turn imposes changes in the overall network of words. New words must be fitted into the system, and unwanted words must be suppressed. Davis and Gaskell (2009) proposed that lexical acquisition consists of two separate processes: an initial rapid familiarization supported by hippocampal learning, followed by a slower process of lexical consolidation that supports the long-term retention of words. As the term is used here, consolidation refers to the strengthening of representations in memory as a result of the integration of new lexical entries into an established semantic network. According to this dual process account, the transfer of knowledge from short-term to long-term retention not only facilitates the access of newly acquired words but also influences the access of existing words as new words are able to compete with similar neighbors.

Tamminen and Gaskell (2008) suggested that allowing time for consolidation should be especially beneficial for production processes as newly established words become integrated into an existing network of words and, thus, gain additional strength. In light of this account, differences in lag time between children in the High Vocabulary and Low Vocabulary groups may reflect variations in the processes that underlie consolidation. If so, we might expect to find a relationship between the degree of lag in comprehension and production and the probability that a word will be either lost or maintained over time. Future research tracking individual words in comprehension and production over longer periods of time is needed to test this hypothesis.

Changes in the structure and operation of the lexicon

In developing a vocabulary, children not only acquire many individual words but also acquire the relations among their meanings. Thus, there are changes in semantic organization that are likely to have consequences for how words are stored and retrieved. As we have suggested, network connectivity and size play an important role in supporting learning and retrieval in terms of both the absolute levels of activation for individual words and the relative strength of neighboring semantic and lexical associates. Given the wealth and organizational efficiency that characterize adults' knowledge of language, it is hardly surprising that, relative to the toddlers in Study 1, adult word learning was accelerated and less susceptible to loss.

For mature speakers, access to words is largely an automatic process. Retrieving words from lexical memory involves procedures that are essentially fast, parallel, and error free (Cohen, Dunbar, & McClelland, 1990). Beginning speakers, by comparison, must invest considerable effort in retrieving the words they wish to say. Once retrieval processes become automatic, we would expect word learners to be faster, closer to ceiling, and less sensitive to small, incremental changes in activation strength as fewer demands are placed on the lexical processing system. This is because the lexicon itself is stable and the operations involved in selecting and retrieving words are well established through practice. Consistent with this prediction, we found that the adults in Study 2 were conspicuously less variable in their performance than the children in Study 1. Interestingly, these results fit with recent observations by Hanania and Smith (2010) of a developmental shift from graded to more categorical
or nongraded processes in tasks of perceived similarity and selective attention. They suggest that this bistability pattern allows the system to shift rapidly from one state of attention to another and represents a more efficient solution to the perceptual task at hand. Similar developments may operate in the lexical domain.

At the neurological level, recent advances in imaging techniques indicate that the development of automaticity in adults involves a reduction in the amount of cortical activity associated with working memory and attentional control (Kelly & Garavan, 2005). Analogous changes are found in the brain activity of infants and children (Durston et al., 2006). For instance, Mills, Plunkett, Prat, and Schafer (2005) used event-related potentials to study changes in hemispheric specialization between 13 and 20 months of age. They found a shift from broad to focal brain activity that was linked to language experience independent of chronological age. In light of these and other findings, Mills and colleagues proposed that neural processes may become more automatic as the amount of brain activity needed to discriminate words decreases and working memory capacity increases. Automaticity, then, may be conceptualized as an emergent property of the language processing system.

What does it mean to know a word?

We have seen that vocabulary learning is not an all-or-nothing affair but rather a messy, noisy, and highly variable process. Dimensions of lexical knowledge develop at different rates, with learners holding varying degrees of word knowledge at any point in time. Our concern in the current study was with the quality of these developmental patterns and not simply with the quantity of learning. The question that naturally arises, however, is whether the fact that learners are able to access a word only in comprehension and not in production means that they do not yet “know” the word. According to Levelt (1989), word knowledge entails the ability to recognize a word when it is heard and to produce the word to express meaning.

An alternative interpretation, however, is that either learners who understand but fail to use a word have not yet developed adequate control over lexical access or their ability to access words has deteriorated. In line with this view, James and MacKay (2007) reported an age-related asymmetry for older adults but for not younger adults. Specifically, they found that older adults (mean age = 73 years) showed greater deficits in production than in recognition, suggesting an increased vulnerability of production processes relative to comprehension. Gershkoff-Stowe (2002) similarly found that young children during the early stages of rapid vocabulary growth often had the correct word in comprehension but still used the wrong word in production. Thus, they appeared to have difficulty in retrieving a known word from lexical memory.1

Limitations and conclusions

The study of comprehension and production processes presents a number of methodological challenges. Different testing methods must be used to tap learners’ knowledge of words, raising the possibility that the observed comprehension–production asymmetry represents an experimental artifact. Although it is difficult to dismiss such an argument, we believe that our results accurately reflect the word learning process for several reasons. First, the finding of a comprehension–production divide is consistent with developmental studies showing differential patterns of acquisition in naturalistic settings (Benedict, 1979; Goldin-Meadow et al., 1976; Harris et al., 1995). Second, we found that adult learners experienced a similar lexical gap despite having advanced attentional and metacognitive skills.

As noted by Bock (1995), recognition and production tasks each present unique problems for the learner. Word comprehension is a form-driven process that involves discriminating a target word from other similar-sounding words. In contrast, word production is a conceptually driven process involving the generation of a word on the basis of semantic input. Thus, differences in the

1 As in the current study, articulatory demands cannot explain the lag in production because children could clearly reproduce the sounds and would imitate the words on request.
computational properties of neural systems that support these basic processes may account for the unique relationship between comprehension and production. An important direction for future study is to obtain converging evidence of the current results by employing alternative methods for testing comprehension and production, for example, eye-tracking and neuroimaging methods (e.g., Tanenhaus, Magnuson, Dahan, & Chambers, 2000; Yu & Smith, 2011). Such studies are needed to confirm the variable patterns of learning found here and to help uncover how learners develop the ability to access both types of information in a rapid and efficient manner.

Two additional limitations warrant consideration. First, although the relatively small number of participants allowed us to look more closely at the word acquisition process by studying individual learning patterns, it also reduced statistical power and, hence, the ability to draw general conclusions. Future studies with greater numbers of children will be required to establish the generality of the findings. Second, training and testing were conducted in a decontextualized setting where little semantic content was available to learners other than the visual properties of the objects. Hence, the range of learning patterns we observed among individual children might not transfer to vocabulary learned in everyday contexts. Although the goal of the current research was to maintain control over the amount of input and practice that learners encountered, future studies might consider increasing the scope and depth of conceptual information related to the target objects as a way of strengthening the semantic connections between words.

Despite these limitations, the findings support several conclusions. First, both novice and mature speakers showed advances in comprehension over production when attempting to learn a new set of words. This result suggests that the asymmetry between comprehension and production is not a developmentally specific phenomenon but rather a reflection of the divergent processes that underlie speaking and understanding. Moreover, the consistency across age groups stresses the importance of linking developmental research on word learning to the adult psycholinguistic literature. Second, 2-year-old word learners did not progress through a series of well-defined stages; rather, their knowledge of words was subject to considerable flux regardless of vocabulary size. This finding strongly suggests that individual words are learned in a graded fashion, with knowledge of a word in different states of knowing at any moment in time.

Most studies of vocabulary growth have considered the acquisition of words in comprehension and production independently. It is highly likely, however, that speaking and listening share a common store of knowledge and are mutually influencing. An important direction for future work is to investigate not only how differences in language use and understanding arise but also the way the two linguistic activities operate together. Such an approach should provide new insights into the nature of the comprehension–production relationship and the dynamic processes that underlie comprehension and production of words.

Acknowledgments

This research was supported by National Science Foundation (NSF) Grant BCS-0841026 to the first author. Special thanks go to the children, parents, and students who participated in the study and to Kaitlyn Crawford, who assisted with the data analyses and preparation of the manuscript.

References


