Using an Artificial Lexicon and Eye Movements to Examine the Development and Microstructure of Lexical Dynamics

James S. Magnuson (magnuson@bcs.rochester.edu)
Delphine Dahan (dahan@bcs.rochester.edu)
Paul D. Allopenna (allopen@bcs.rochester.edu)
Michael K. Tanenhaus (mtan@bcs.rochester.edu)
Richard N. Aslin (aslin@cvs.rochester.edu)
Department of Brain and Cognitive Sciences
University of Rochester, Meliora Hall, Rochester, NY 14627 USA

Abstract

It is well known that the time course of lexical access is shaped by the number and nature of potential competitor items in the lexicon. While research has outlined the *macrostructure* of lexical processing (e.g., that during spoken word recognition, lexical candidates similar to the input are activated and compete for recognition), many questions remain about the *microstructure* (how exactly is the competitor set defined?) and *dynamics* (what is the time course of lexical competition?) of lexical processing, as well as their *development* as words are learned. Here, we begin to address these issues with a study in which participants learned to recognize words from a lexicon of novel names associated with novel shapes. Each item in the lexicon (e.g., /pibo/) had two potential competitors (e.g., /pibu/ and /dibo/). Half of the words were presented more frequently than the other half during training. This allowed us to examine the development of competition effects with experience. An eye tracker provided an on-line measure of the items being considered for recognition. The results indicate that lexical competition effects among newly-learned items develop quickly.

Introduction

In recent years, issues of lexical representation and process have taken on an increasingly central role in models of language comprehension. Lexical representations are now viewed as providing much of the syntactic, semantic, and pragmatic knowledge necessary for the early stages of parsing and interpretation (e.g., MacDonald, Pearlmutter & Seidenberg, 1994; Tanenhaus & Trueswell, 1995). In the domain of spoken language, lexical knowledge is implicated in aspects of speech recognition that were often previously viewed as pre-lexical (Andruski, Blumstein, & Burton, 1994; Marslen-Wilson & Warren, 1994). Thus, having a detailed understanding of lexical processing is essential for the broader goal of developing theories of language comprehension and development.

Three key findings indicate that during spoken word recognition listeners evaluate the unfolding input against an *activated* set of lexical *candidates* which *compete* for recognition. First, many spoken words, especially polysyllabic content words, are often recognized before the end of the word, as assessed by a variety of speeded reaction time measures (e.g., Marslen-Wilson, 1987). Second, the time-course of recognition depends, in part, on constraints provided by sentential context (Zwitserlood, 1988). Third, recognition time is contingent in that it depends not only on the properties of the input and the target word (e.g., its frequency, phonological structure, etc.), but also on its similarity to phonetically similar lexical candidates (Andruski et al, 1994; Luce, Pisoni, & Goldinger, 1990; Marslen-Wilson, 1987; 1993).

These results provide a clear picture of what Marslen-Wilson (1993) has termed the *macrostructure* of spoken word recognition. However, numerous questions remain about the *microstructure* of the dynamics of on-line lexical processing (what determines the nature of the competitor set and the time course of competition effects), as well as the development of these dynamics as words are learned. There have been few studies of the development of lexical *processing*. For example, Charles-Luce and Luce (1990) have described neighborhoods in children's lexicons (based on the number of similar lexical items for each lexical item). An important exception is work by Swingley (1997), who examined on-line onset cohort effects (parallel consideration of lexical items which share onsets) in 18- and 24-month-olds. While this work supports the hypothesis that processing is incremental and competitive even when children's vocabularies are small, it does not address how well words must be learned before such effects become apparent (Swingley's stimuli were words the children in his study already knew). That is, is incremental processing the natural mode of spoken word recognition, or does it emerge only when lexical items are well-learned?

Here, we begin to address these issues with a study in which participants learned to associate sixteen novel names with sixteen novel shapes. Each of the sixteen names in the artificial lexicon (e.g., /pibo/) had an onset cohort competitor (e.g., /pibu/) and a rhyme competitor (e.g., /dibo/). This allowed us to examine specific competition effects which differentiate two major classes of spoken word recognition models (which will be discussed in more detail in the next section). In addition, half of the words were presented with high frequency during training. This allowed us to study differences in activation due to differing amounts of experience during training.

In order to measure competition effects on-line, we used a recently-developed eye tracking methodology (the "visual world paradigm"; see Tanenhaus and Spivey-Knowlton, 1996, for methodological details) which has allowed extremely fine-grained measurements of the time course of competition during various aspects of language

comprehension (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). We will now review recent studies using this methodology which demonstrate its usefulness for observing incremental consideration of lexical competitors as a spoken word unfolds over time.

The Visual World Paradigm

In the visual world paradigm, participants wear a light-weight, head-mounted eye tracker as they follow instructions to interact with objects in the visual world (e.g., "pick up the beaker; now put it next to the square"). The paradigm has several advantages over conventional psycholinguistic tasks. For example, unlike a task like lexical decision, no metalinguistic judgment is called for; participants behave naturally, and incidental fixations are recorded. Further, given a task which requires visual guidance (e.g., picking up objects), eye movements are closely time locked to the speech stream (note that using a task for which fixations are functionally relevant avoids the problems of interpretation discussed by Viviani, 1990; see Allopenna, Magnuson, & Tanenhaus, in press, for more discussion).

For example, Spivey-Knowlton and Tanenhaus (submitted) found that, given a display containing two cohorts (words sharing onsets, such as "candy" and "candle") as well as phonetically unrelated distractors, subjects were initially equally likely to fixate "candy" and "candle" (and significantly more likely to fixate those items than distractors) when told to pick up the candy. After the onset of the final vowel of "candy," the probability of fixating "candle" quickly fell to zero.

Allopenna et al. (in press) used the paradigm to test differing predictions of alignment models such as Cohort (e.g., Marslen-Wilson, 1987) and continuous activation models such as TRACE (McClelland & Elman, 1986) and Shortlist (Norris, 1994). In the Cohort model, as a word unfolds over time, potential matches to the input are activated and compete for recognition. Given the word "beaker", early on, all items beginning with /b/ will become slightly active (i.e., will form the current "cohort"). As more information is heard, items which are inconsistent with the input will be strongly inhibited (or eliminated). This continues until the lexical item which best matches the input is identified. Thus, competition is predicted among words which share onsets (e.g., "beaker" and "beetle", which we will refer to as "onset cohort competitors"), but competition among rhymes (e.g., "beaker" and "speaker") is held to be extremely unlikely, since rhymes are likely to be excluded from the cohort early on.

In contrast, continuous activation models like TRACE predict activation of rhymes due to their overall similarity. In TRACE, units at the word level are connected to input units which represent phonetic features. Similarity between a word unit and the input late in a word will still result in activation of the word unit. Rhyme activation is predicted to be relatively lower than that of cohort competitors because of their initial mismatch, but substantial nonetheless.

While cohort effects are well-established (e.g., Grosjean, 1980; Marslen-Wilson, 1989; Tyler, 1984; Zwitserlood, 1989), rhyme effects have proven more elusive (e.g., Zwitserlood, 1996). Although there have been recent reports of rhyme effects in cross-modal and auditory-auditory priming (Connine, Blasko & Titone, 1993; Andruski et al., 1994; Marslen-Wilson, Moss, & van Halen, 1996), the effects have involved stimuli differing by only one or two phonetic features. This suggests that only very small mismatches are tolerated, which is consistent with a probabilistic form of an alignment model (cf. Marslen-Wilson et al., 1996).

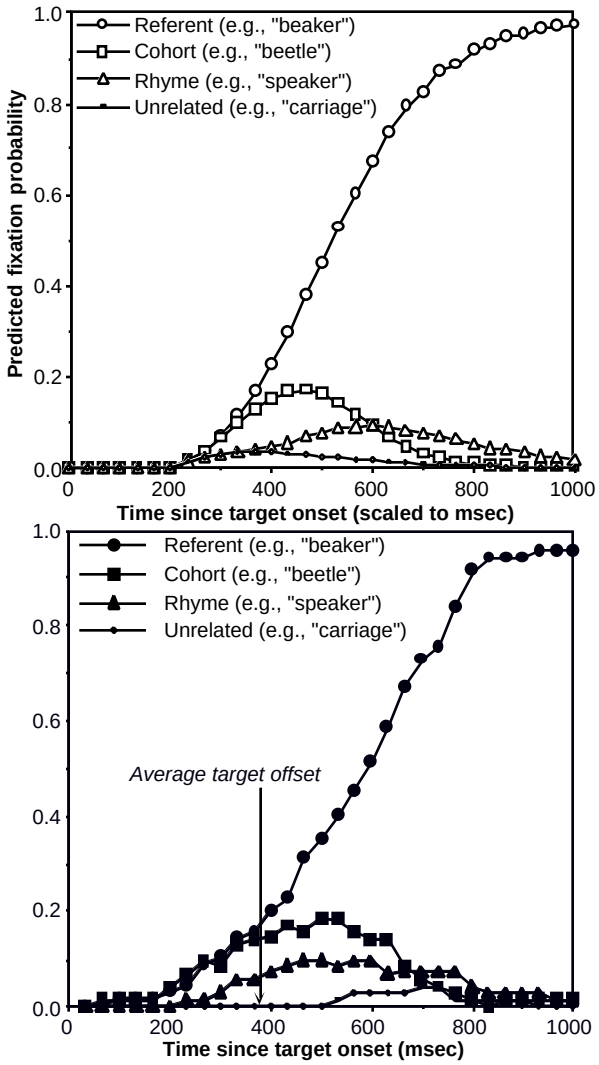

Figure 1: TRACE activations converted to predicted fixation probabilities (top panel) and observed probabilities of fixating a target referent, a cohort member, a rhyme, and an unrelated object from Allopenna et al. (in press).

However, using the visual world paradigm (with displays containing, e.g., a *beaker*, a *beetle*, a *speaker*, and a *carriage*, and instructions like "pick up the beaker"), Allopenna et al. (in press) found evidence for rhyme competition even with rhyme stimuli that differed by more than two features. In Figure 1, simulation and experimental data from Allopenna et al. are presented. As predicted by continuous activation models (here, TRACE activations converted to fixation probabilities using a variant of the Luce choice rule; see

Allopenna et al. for details), there is evidence of substantial rhyme activation beginning shortly after the beginning of the vowel, and reaching a peak activation (or fixation probability) somewhat less than that of the cohort.

In very simple tasks, participants require approximately 150 msec to plan and launch a saccade (e.g., Matin, Shao, & Boff, 1993). Allowing for this planning time, it is clear that the earliest eye movements are being planned approximately 100 msec after target onset. Thus, the visual world paradigm is able to detect very subtle effects, and so provides an extremely sensitive measure of processing which is time-locked to the speech stream, and which requires only natural actions in response to spoken instructions.

Learning and the Visual World Paradigm

Previous studies have used artificial words to study word segmentation (Saffran, Newport, & Aslin, 1996) or artificial words and grammars to study syntactic development (e.g., Braine, 1963; Morgan, Meier, & Newport, 1987). Our goal here was to examine the development of lexical dynamics as words are learned. We used a lexicon of novel names because this allowed us to have control over factors such as the detailed nature of the competitor set and frequency.[1] In addition, since the names must be learned, we can observe whether competition among phonologically similar items emerges early in learning, or if such effects depend on the items being very well-learned.

We gave participants about 80 minutes of training to associate 16 nonsense shapes with 16 bisyllabic novel names. Each name had one onset competitor and one rhyme competitor. Neighborhood was held constant (all words had neighborhoods of three; the word itself -- e.g., /pibo/ -- plus an onset competitor and a rhyme -- e.g., /pibu/ and /dibo/). Word frequency was manipulated by presenting eight of the words with relatively high frequency, and the other eight with relatively low frequency. In addition, four of the high-frequency items had low-frequency competitors, and four had high-frequency competitors. The same was true for the low-frequency items. These manipulations were introduced to allow us to examine whether any competition effects that might emerge would be modulated by frequency.

Method

Participants Five students at the University of Rochester were paid for their participation. All were native speakers of English with normal or corrected-to-normal vision.

Materials The visual stimuli were simple patterns, formed by filling eight randomly-chosen, contiguous cells of a four-by-four grid (see Figure 2). 10,000 such randomly-generated patterns were randomly ordered, and sixteen were selected from the beginning of the set (with two items replaced due to visual similarity with other items).

[1] One potential concern with such a learning study is the intrusion of participants' English lexicons. This is an issue we defer for future research, once we have established that lexical processing effects can be observed with newly-learned items.

The phonological materials consisted of sixteen bisyllabic nonsense words. The sixteen words comprised four four-word sets, such as /pibo/, /pibu/, /dibo/, and /dibu/. Note that for each word, there is an onset competitor which differs only in the final vowel, a rhyme, and a relatively dissimilar item (differing by two phonemes, which would not qualify it as a neighbor using the most standard definition of a word differing by a single phoneme). A small set of phonemes was selected in order to achieve consistent similarity within and between sets. The consonants /p/, /b/, /t/, and /d/ were chosen because they are among the most phonetically similar stop consonants. In each set, rhymes differed by two phonetic features (place and voicing) in the first phoneme. Transitional probabilities were controlled such that all phonemes and combinations of phonemes were equally predictive at each position or combination of positions.

The auditory stimuli were produced by a male, native speaker of English in a sentence context ("click on the ____"). The average duration of the target words was 496 msec. The stimuli were recorded to tape, and then digitized using the standard analog/digital devices on an Apple Macintosh 8500 at 16 bit, 44.1 kHz. The stimuli were converted to 8 bit, 11.127 kHz (SoundEdit format) in order to be used with the experimental control software, PsyScope 1.2 (Cohen, MacWhinney, Flatt & Provost, 1993).

Procedure Participants came to the lab for two 2-hour sessions on consecutive days. Each day consisted of seven training sessions with feedback and a testing session without feedback. We tracked eye movements during the test.

Participants were seated at a comfortable distance from the experimental control computer (an Apple Macintosh 7200 PowerPC). The structure of the training sessions was as follows. First, a fixation cross would appear on the screen. The participant had to click on the cross to begin the trial. After 500 msec, either two shapes (in the first three training sessions) or four shapes (in the rest of the training sessions and the tests) would appear. If only two shapes were presented, they appeared approximately 1.5 degrees to the left and right of the fixation cross. When four shapes were presented, two would also appear about 1.5 degrees above and below the fixation cross (see Figure 2).

Participants would hear the instruction, "Look at the cross" through headphones 750 msec after the objects appeared. At this point, participants would fixate the cross and click on it. Participants were instructed at the beginning of the session that they needed to fixate the cross until they heard the next instruction. Then, 500 msec after clicking on

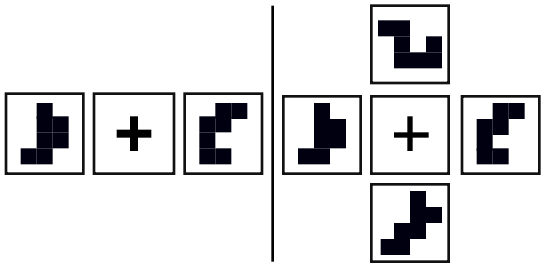

Figure 2: The left panel shows a possible display in 2AFC training; the right panel shows a possible 4AFC display.

the cross, an instruction to click on one of the items (with the computer's mouse) would be presented (e.g., "Click on the pibu").

When participants responded by clicking on one of the items, or at the end of 15 seconds, all of the items would disappear except for the shape that was actually named. The shape's name was repeated (e.g., "pibu") 500 msec later, so that the feedback was the same on every trial, regardless of whether the participant clicked on the correct shape or not. The object disappeared 500 msec later, and the subject would click on the cross to begin the next trial. The testing session was identical to the four-item training, except that no feedback was given.

Shapes were randomly mapped to names. Two random mappings were used, with two or three subjects trained on each of the mappings. Half the items were high- and half were low-frequency. In addition, half of the eight high-frequency items had low-frequency competitors (e.g., /pibo/ and /dibu/ might be high frequency, and /pibu/ and /dibo/ would be low frequency). The same was true of the low frequency items.

Each training session consisted of 64 trials. High frequency names appeared 7 times per session, and low-frequency names appeared once per session. In addition to their use as targets, shapes were used as visual distractors. Each item appeared 4 times as a visual distractor. Thus, each high-frequency item appeared 42 times as a target and each low-frequency item appeared 6 times as a target during the six training sessions, and each shape appeared 24 times as a visual distractor. Within training, distractors were randomly assigned to each trial. Items were not allowed to appear as the target on consecutive trials.

The tests consisted of 96 trials. Each item appeared in six trials: one with its onset competitor and two unrelated items, one with its rhyme competitor and two unrelated items, and four with three unrelated items. Unrelated items were matched in frequency to the target's competitors.

We tracked eye-movements with an Applied Scientific Laboratories (E4000) eye tracker. Two cameras mounted on a lightweight helmet provide the input to the tracker. The eye camera provides an infrared image of the eye. The center of the pupil and the first Purkinje corneal reflection are tracked to determine the position of the eye relative to the head. Accuracy is better than 1 degree of arc, with virtually unrestricted head and body movements. A scene camera is aligned with the participant's line of sight. A calibration procedure allows software running on a PC to superimpose crosshairs showing the point of gaze on a HI-8 video tape record of the scene camera. The scene camera samples at a rate of 30 frames per second, and each frame is stamped with a time code. The auditory stimuli were presented binaurally through headphones using the standard digital-to-analog devices provided with the experimental control computer. Audio connections between the computer and HI-8 VCR provided an audio record of each trial.

Results

Participants were able to achieve high levels of accuracy relatively quickly. The accuracy results for the fourteen training sessions and the two test sessions are shown in Table 1.

Table 1: Accuracy in training and testing sessions.

Day 1			Day 2		
Type	High	Low	Type	High	Low
2 AFC	.70	.59	4 AFC	.94	.75
2 AFC	.83	.44	4 AFC	.96	.83
2 AFC	.90	.59	4 AFC	.95	.90
2 AFC	.95	.75	4 AFC	.94	.85
4 AFC	.90	.64	4 AFC	.99	.88
4 AFC	.94	.68	4 AFC	.99	.95
4 AFC	.94	.70	4 AFC	.96	.93
Test	.95	.76	Test	.98	.98

Table 1. On the first day, participants clearly learned the high-frequency items better than the low-frequency items. Participants showed slight drops in accuracy when the task was changed to four-alternative forced choice, and again when feedback was discontinued in the test. Note that participants were close to ceiling for high-frequency items in the first test, but did not reach ceiling for low-frequency items until the end of the second day.

Figure 3 shows the probability of making eye-movements to cohort, rhyme and unrelated distractors, averaged across all frequency and competitor conditions in the test at the end of day 2. Note that the time axis extends further than that in Figure 1, and that the probability of fixating the target shape is still relatively low even 1500 msec after the onset of the target name. Two factors underlie this. First, the stimuli are longer than average bisyllabic words because of their CVCV structure. Second, although subjects are at ceiling on high- and low-frequency items in the second test, they are still not as confident in their responses as they would be for real words, and make more eye movements after word onset than participants in Allopenna et al. (in press). However,

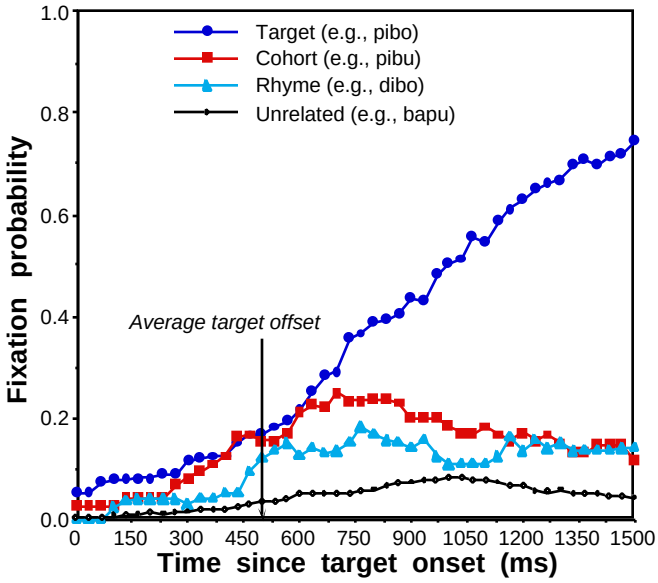

Figure 3: Cohort and rhyme effects averaged across all conditions in the test at the end of the second day.

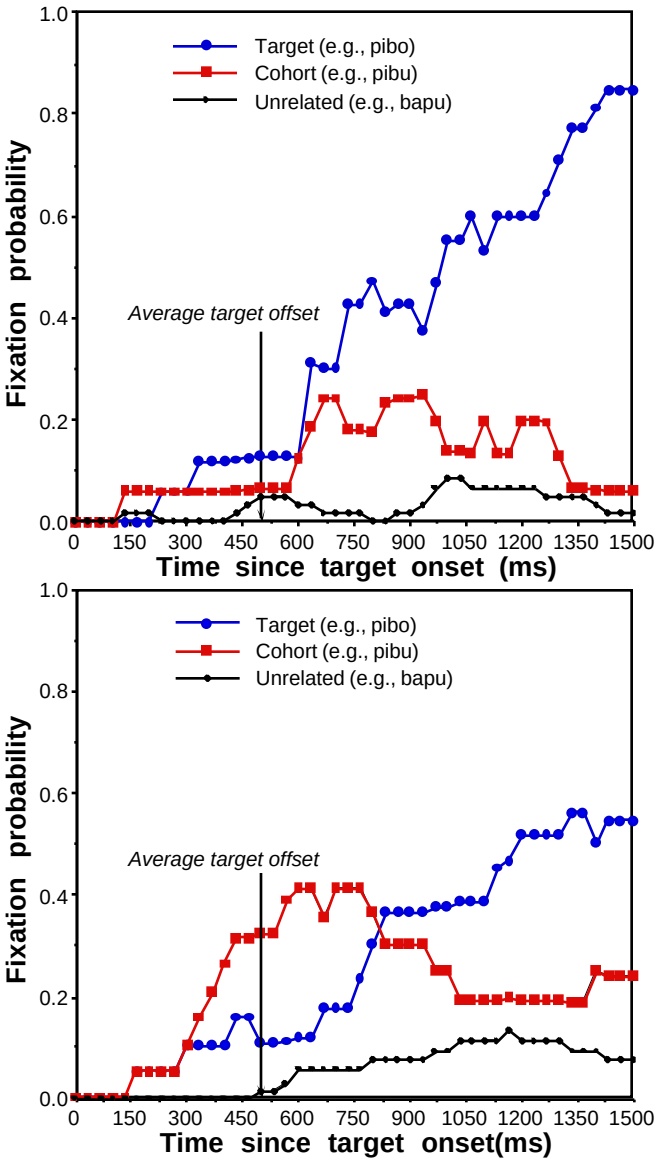

Figure 4: Cohort effects modulated by frequency in the test on day 2. In the upper panel, both the target and the cohort are high-frequency. In the lower panel, the target is low-frequency and the cohort is high-frequency.

recall that it takes approximately 150 to 200 msec to plan and initiate a saccade in a simple task. This means that at least the fixation probabilities observed in the first several hundred milliseconds are time-locked with the spoken input.

As can be seen in Figure 3, there were cohort and rhyme effects that emerged as a function of similarity with the target. The cohort and target probabilities separate together from the unrelated baseline. After a slight delay, the probability of fixating the rhyme separates from baseline.

The competitor effects were also modulated by frequency. Two of the four combinations of target and competitor frequency are shown in Figure 4. In the upper panel, the cohort condition in which both the target and the cohort are high frequency is shown. The pattern is similar to the

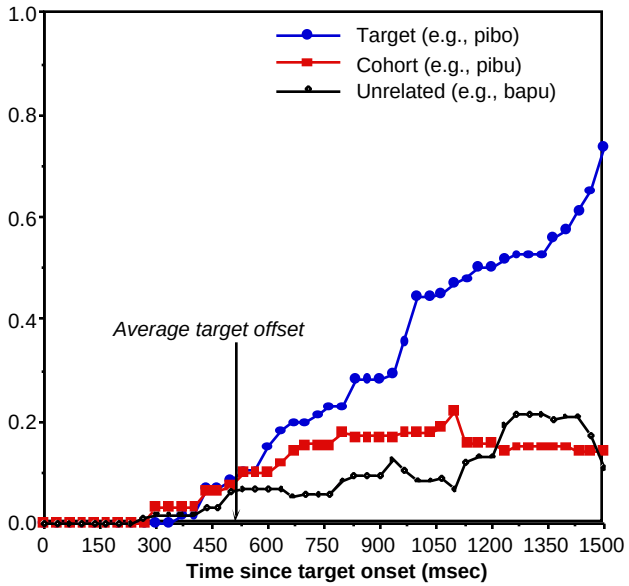

Figure 5: The cohort effect after only one day of training.

overall pattern shown in Figure 3. In the lower panel, the cohort condition in which the cohort is high frequency but the target is low frequency is shown. There is a striking effect of frequency. Participants were much more likely to fixate the cohort early in the word when the input was consistent with the high frequency cohorts and targets. Although they are not shown here, a slight advantage is seen for the target item in the condition where it is high frequency and the cohort is low frequency, and no advantage is seen when both items are low frequency. Similar effects are seen for rhyme competitors, but are not shown here.

Discussion

The results indicate that with relatively little training, participants form representations of novel words which begin to approximate those for real words. Eye-movement measures over time suggested that participants were considering onset competitor items soon after word onset. Analyses by frequency condition showed that this effect was modulated by the frequency manipulation. Thus, the 98 exposures to high-frequency items and 14 to low-frequency items during training were sufficient to begin to approximate the lexical dynamics observed with real words. In fact, after just 49 exposures to high-frequency items and 7 exposures to low-frequency items on the first day of training, cohort effects were already present (see Figure 5), but there were no rhyme effects.

The absence of a rhyme effect on day 1 provides an interesting parallel to Swingley's (1997) failure to find rhyme effects with 24-month old children (although he did find onset cohort effects). It may be that the absence of a rhyme effect is diagnostic of a stage in which lexical representations are not well-enough established for subtle similarity effects to emerge. Future work will explore the differences in the amount of learning necessary to establish various lexical processing effects.

This work makes three contributions to the literature: (1) it shows that lexical competition effects can develop quite quickly; (2) it reinforces previous demonstrations that on-line language comprehension is incremental; and (3) the results indicate that this paradigm can be extended to more complex issues of lexical structure and lexical acquisition. Overall, the results show that incremental consideration of multiple alternatives in parallel does not depend on highly-learned lexical items; it is the natural mode for spoken word recognition.

Acknowledgments

Supported by NIH HD27206 to MKT, NSF SBR-9729095 to MKT and RNA, and an NSF Graduate Research Fellowship to JSM.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (in press). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language.*

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, 52, 163-187.

Braine, M. D. S. (1963). On learning the grammatical order of words. *Psychological Review*, 70, 323-348.

Charles-Luce, J. & Luce, P.A. (1990). Similarity neighborhoods of words in young children's lexicons. *Journal of Child Language*, 17, 205-215.

Cohen J.D., MacWhinney B., Flatt M. & Provost J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments & Computers*, 25(2), 257-271.

Connine, C. M., Blasko, D. G., and Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language*, 32, 193-210.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, 28, 267-283.

Luce, P. A., Pisoni, D. B., and Goldinger, S. D. (1990). Similarity neighborhoods of spoken words. In G. T. M. Altmann (Ed.), *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*. Cambridge, MA: MIT.

MacDonald, M. C, Pearlmutter, N.J., & Seidenberg, M.S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676-703.

Marslen-Wilson, W. (1987). Functional parallelism in spoken word recognition. *Cognition*, 25, 71-102.

Marslen-Wilson, W. (1993). Issues of process and representation in lexical access. In G.T.M. Altmann & R. Shillcock (Eds.), *Cognitive Models of Speech Processing: The Second Sperlonga Meeting.* Erlbaum.

Marslen-Wilson, W., Moss, H.D., & van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 1376-1392.

Marslen-Wilson, W. & Warren, P. (1994). Levels of perceptual representation and process in lexical access: words, phonemes, and features. *Psychological Review*, 101, 653-675.

Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 576-585.

Matin, E., Shao, K. C., and Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, 53, 372-380.

McClelland, J.L., & Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1-86.

Morgan, J. L., Meier, R. P., and Newport, E. L. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology*, 19, 498-550.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52, 189-234.

Saffran, J. R., Newport, E. L., and Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35, 606-621.

Spivey-Knowlton, M. J., & Tanenhaus, M. K. (submitted). Integration of visual and linguistic information in resolving temporary lexical and structural ambiguities.

Swingley, D. (1997). *Word Recognition and Representation in Young Children*. Unpublished Ph.D. thesis, Stanford University.

Tanenhaus, M.K. & Spivey-Knowlton, M.J. (1996). Eye-tracking. In F. Grosjean & U. Frauenfelder (Eds). *Special Issue of Language and Cognitive Processes: A Guide to Spoken Word Recognition Paradigms*, 11, 583-588.

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

Tanenhaus, M.K. & Trueswell, J. C. (1995). Sentence comprehension. In J. L. Miller and P. D. Eimas (Eds.), *Handbook of Perception and Cognition Volume 11: Speech, Language, And Communication*. San Diego: Academic Press.

Tyler, L. (1984). The structure of the initial cohort: Evidence from gating. *Perception & Psychophysics*, 36(417-427).

Viviani, P. (1990). Eye movements in visual search: Cognitive, perceptual, and motor control aspects. In E. Kowler (Ed.), *Eye Movements and Their Role in Visual and Cognitive Processes. Reviews of Oculomotor Research V4.* Amsterdam: Elsevier.

Zwitserlood, P. (1989). The locus of the effects of the sentential-semantic context in spoken word processing. *Cognition*, 32, 25-64.

Zwitserlood, P. (1996). Form priming. In F. Grosjean and U. Frauenfelder (Eds). *Special Issue of Language and Cognitive Processes: A Guide to Spoken Word Recognition Paradigms*, 11, 589-596.