Journal of Experimental Psychology:
Learning, Memory, and Cognition
2004, Vol. 30, No. 2, 498–513

Copyright 2004 by the American Psychological Association, Inc.
0278-7393/04/$12.00 DOI: 10.1037/0278-7393.30.2.498

Continuous Mapping From Sound to Meaning in Spoken-Language Comprehension: Immediate Effects of Verb-Based Thematic Constraints

Delphine Dahan
Max Planck Institute for Psycholinguistics

Michael K. Tanenhaus
University of Rochester

The authors used 2 "visual-world" eye-tracking experiments to examine lexical access using Dutch constructions in which the verb did or did not place semantic constraints on its subsequent subject noun phrase. In Experiment 1, fixations to the picture of a cohort competitor (overlapping with the onset of the referent's name, the subject) did not differ from fixations to a distractor in the constraining-verb condition. In Experiment 2, cross-splicing introduced phonetic information that temporarily biased the input toward the cohort competitor. Fixations to the cohort competitor temporarily increased in both the neutral and constraining conditions. These results favor models in which mapping from the input onto meaning is continuous over models in which contextual effects follow access of an initial form-based competitor set.

As people listen to spoken utterances, they rapidly map the incoming acoustic input onto meaning-based representations that take into account the speech input and the relevant linguistic and nonlinguistic context (e.g., Marslen-Wilson, 1975, 1987, 1989; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). At any point in the unfolding speech stream, numerous words will be at least partially consistent with the speech input. For example, the initial sounds /kæ/ in *candy* will partially match multiple lexical candidates, including *can, candy, candle,* and *candid,* among others. However, only a subset of the candidate words is likely to be consistent with the listener's developing interpretation at the point where the input is encountered. For example, in an utterance beginning "Thomas is not allowed to snack on . . . ," *candy* is a better fit to the context than *can, candle,* or *candid,* because only *candy* is a plausible theme of the verb *snack on.* The research reported here addresses the time course with which acoustic input is mapped onto semantic representations by examining (a) how verb-based semantic constraints influence lexical access and (b) how these constraints affect the processing of subsequent phonetic information.

Delphine Dahan, Max Planck Institute for Psycholinguistics; Michael K. Tanenhaus, Department of Brain and Cognitive Sciences, University of Rochester.

This work was partially supported by National Institute on Deafness and Other Communication Disorders Grant DC-005071 and The National Institute of Child Health and Human Development Grant HD-27206. We thank Tau van Dijck for his tremendous help throughout the preparation and running of these experiments, Manon van Laer for her assistance with testing the subjects in the eye-tracking experiments, and Margret van Beuningen and Michiel Seevinck for preparing and running the norming experiments. We also thank Gerry Altmann and Maryellen MacDonald for helpful comments on drafts of this article.

Correspondence concerning this article should be addressed to Delphine Dahan, who is now at the Department of Psychology, University of Pennsylvania, 3401 Walnut Street, Suite 400A, Philadelphia, PA 19104. E-mail: dahan@psych.upenn.edu

According to an influential perspective perhaps most clearly articulated by Marslen-Wilson (1987, 1989), spoken-word recognition can be divided into two partially overlapping subprocesses: *access* and *selection.* During access, the spoken input is mapped onto sound-form lexical representations, activating a set of lexical candidates. Access makes available lexically specific syntactic and semantic information about each candidate. A single best-fitting candidate is then selected, taking into account goodness of fit with the input and the context, and integrated with the current representation. The access–selection framework assumes a temporal window during which the activation of the syntactic and semantic components of a lexical candidate is determined solely on the basis of its phonetic match to the input, without influence from, or integration with, context. Under this view, context has a delayed influence on the recognition of a spoken word in spite of being available prior to sensory information about the word itself. Although Marslen-Wilson and colleagues (e.g., Gaskell & Marslen-Wilson, 1997, 2002) have subsequently developed a distributed model that blurs the distinction between access and selection, the access–selection framework remains an influential perspective in spoken-word recognition (e.g., Connine, Blasko, & Wang, 1994; Miller & Eimas, 1995).

Perhaps the most striking support for context-independent, form-based access comes from a study by Zwitserlood (1989). Zwitserlood used a cross-modal semantic priming paradigm in which gated fragments of Dutch polysyllabic words, such as *kapitein* [captain], were followed by a visually presented target word for lexical decision. The target word was semantically related either to the spoken word itself (e.g., *schip* [ship], semantically related to *kapitein*) or to a cohort competitor sharing the same initial sounds (e.g., *geld* [money], semantically related to *kapitaal* [capital]). When gated fragments that were still consistent with both the actual word and its cohort competitor were presented in biasing sentential contexts, targets related to the contextually incongruent cohort competitor showed as much priming as did targets related to the contextually congruent word. However, priming to an incongruent cohort competitor declined more rapidly in the biasing sentential context compared with a neutral context. The decline was present for gated fragments that did not yet include phonetic information that distinguished between the actual word and its competitor. Thus, the data pattern was fully consistent with the idea that initial form-based access made available lexical candidates and their semantic properties (allowing facilitation of a response to a semantically related word), with context affecting selection but not access (see Connine [1987] and Connine et al. [1994] for similar conclusions).

Zwitserlood's (1989) results were also consistent with a body of cross-modal semantic priming evidence from previous studies with fully ambiguous words, such as *bank.* When targets were presented immediately at the offset of the ambiguous word, priming was found to associates of either sense of the word, regardless of context (Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982; Swinney, 1979; Tanenhaus, Leiman, & Seidenberg, 1979). Other evidence consistent with context-independent access is reviewed in Marslen-Wilson (1987, 1989).

More recent results, however, suggest that the access–selection framework may need to be revised in subtle but theoretically important ways. First, recent research demonstrates that the system is much more tolerant of phonetic mismatches than the access–selection framework assumes it is. Lexical candidates that mismatch the input at onset by several phonetically distinctive features are nonetheless partially activated as a word unfolds, even though fully matching lexical candidates are already activated (Allopenna, Magnuson, & Tanenhaus, 1998; Connine, Blasko, & Titone, 1993). This result weakens the "propose and dispose" aspect of access, because it becomes more difficult to isolate a word–initial candidate set.

There is also reason to question whether context-based selection and integration lag appreciably behind access. In studies using cross-modal semantic priming, Tabossi and colleagues (Tabossi, 1988; Tabossi, Colombo, & Job, 1987; Tabossi & Zardon, 1993) found little or no priming to associates of the less frequent sense of an ambiguous word in contexts that provide a strong bias for the dominant sense, even before the ambiguous word has been fully heard. Given that one sees effects of lexical frequency very early in lexical processing (e.g., Dahan, Magnuson, & Tanenhaus, 2001; Marslen-Wilson, 1993), this suggests that selection is rapid enough for integration to be completed before the less frequent sense becomes highly activated (for similar results in reading, see Duffy, Morris, & Rayner, 1988; Rayner, 1998; for reviews, see Lucas, 1999; Simpson, 1994).

Recent results from studies monitoring event-related brain potentials also demonstrate that the semantic and discourse-based fit of words can be established extremely rapidly, on the basis of even partial acoustic information and with no apparent delay (Connolly & Phillips, 1994; van Berkum, Zwitserlood, Hagoort, & Brown, 2003; van den Brink, Brown, & Hagoort, 2001; Van Petten, Coulson, Rubin, Plante, & Parks, 1999).

In light of the emerging evidence we have reviewed, it is important to directly test the central assumption underlying the access–selection framework, namely that there is a brief but measurable temporal interval during which lexically based syntactic and semantic information is made available solely by form-driven access processes. The alternative is that mapping from the input to meaning is a continuous process with emerging representations that are continuously updated using multiple information sources. In such a system, there is no discrete candidate set, no discrete point in time at which lexical selection is completed, and the system remains continuously tuned to the input (e.g., Gaskell & Marslen-Wilson, 1997, 2002).

In order to contrast the access–selection and continuous-mapping models, we examined the processing of lexical competitors in sentential contexts using naturally occurring Dutch constructions in which the verb precedes its subject noun phrase. The subject noun phrase followed either a main verb that placed strong semantic constraints on its subject (e.g., *Nog nooit klom een bok zo hoog* [Never before climbed a goat so high]) or an auxiliary or modal verb that did not place such constraints on its subject (e.g., *Nog nooit is een bok zo hoog geklommen* [Never before has a goat climbed so high]).

We manipulated verb-based semantic constraints because the recognition of a verb makes available its combinatory syntactic and semantic constraints (e.g., Altmann, 1999; Boland, 1997; Boland, Tanenhaus, Garnsey, & Carlson, 1995; Ferretti, McRae, & Hatherell, 2001; McRae, Ferretti, & Amyote, 1997; Tanenhaus, Garnsey, & Boland, 1990). These constraints include information about the types of complements that occur with a verb and the semantic (thematic) constraints that apply to these complements. Following McRae et al. (1997), we view thematic constraints as verb-specific concepts associated with the arguments of a verb, rather than as coarse-grained selectional restrictions, such as animacy. We should note, though, that our materials did not allow us to distinguish between these alternatives. Crucially, verb-based constraints are available rapidly enough to influence how a subsequent noun phrase is parsed when the noun phrase would otherwise be consistent with two alternative syntactic structures (Garnsey, Pearlmutter, Myers, & Lotocky, 1997; Trueswell & Kim, 1998; Trueswell, Tanenhaus, & Kello, 1993).

Perhaps the most convincing evidence that a verb makes available semantic constraints about its upcoming arguments comes from recent eye-tracking work by Altmann and Kamide (1999). Altmann and Kamide presented participants with an auditory sentence and, concurrently, a visual scene containing, for example, a boy, a cake, a toy train, a toy car, and a ball. In nonconstraining contexts, such as "The boy will move . . ." multiple pictures were possible themes of the verb *move,* including the actual theme *cake,* whereas in constraining contexts, for example, "The boy will eat . . ." only one object (the cake) satisfied the constraints of the verb. For the constraining verbs, participants made anticipatory eye movements to the thematically congruent referent beginning at the verb and preceding the onset of the noun, demonstrating that verb-based information is accessed and integrated quickly enough to allow listeners to anticipate and evaluate likely referents (see also Boland, 2001; Kako & Trueswell, 2000; Kamide, Altmann, & Haywood, 2003). Crucially, for our purposes, these results indicate that verb-based information that predicts semantic properties of the upcoming subject is likely to be available when the upcoming noun is encountered.

We examined the processing of a spoken noun following a semantically constraining or neutral verb by monitoring listeners' eye movements to pictured objects displayed on a computer screen as they heard the sentence. Their task was to indicate, by clicking on it, which of the four objects was mentioned in the sentence. The target object was mentioned as the subject noun following the constraining or neutral verb. Before turning to the details of the experiments, we consider some methodological issues involved in our choice of paradigm.

Advantages of the Visual-World Paradigm

We used the "visual-world" eye-tracking paradigm because it provides a fine-grained measure of the time course of lexical processing in continuous speech that is closely time-locked to the input. The average minimal latency for planning and launching a saccade in a number of tasks has been estimated to be approximately 150–200 ms (e.g., Fischer, 1992; Hallet, 1986; Pollatsek, Rayner, & Collins, 1984; Rayner, Slowiaczek, Clifton, & Bertera, 1983; Saslow, 1967). As found in previous research with neutral contexts (e.g., Allopenna et al., 1998), the probability of fixating pictures with cohort competitor names (e.g., the picture of a beetle) begins to increase about 200 ms after the onset of the target word (e.g., the word *beaker*). Fixations to the pictured referent begin to diverge from those to a cohort competitor within 200–300 ms after disambiguating phonetic information is first encountered. Moreover, the proportion of fixations at a point in time can be closely mapped onto underlying lexical activation using a simple linking hypothesis in which the activation of lexical candidates is affected by the entire lexicon, but the response strengths for the potential referents, in other words, those that are pictured, are evaluated using the Luce choice rule (see Allopenna et al., 1998; Dahan, Magnuson, & Tanenhaus, 2001; Magnuson, Tanenhaus, Aslin, & Dahan, 2003; Tanenhaus, Magnuson, Dahan, & Chambers, 2000).

Fixations generated as a spoken word is processed also capture transient and fine-grained modulations in lexical activation that other online paradigms have sometimes failed to reveal. For example, Allopenna et al. (1998) found clear evidence for activation of words that rhymed with a target word (e.g., activation of *handle* when the target word was *candle*), even when the onset of the rhyme competitor differed from the target by several phonetic features. McMurray, Tanenhaus, and Aslin (2002) showed gradient effects of within-category voice onset time. The proportion of fixations to a competitor picture, for example, a peach, was a linear function of voice onset time to the target word, for example, *beach.* Dahan, Magnuson, Tanenhaus, and Hogan (2001) manipulated the acoustic realization of the name of the referent picture by splicing the final portion of the word onto the initial sounds of a competitor word. For example, an altered version of the word *net* was created by splicing the final *t* from a token of *net* onto the initial portion—including the vowel—of the competitor word *neck.* Because the realization of a vowel is subject to articulatory influence from its surrounding consonants, the vowel /ε/ was predicted to partly encode the identity of its subsequent consonant /k/, temporarily favoring a *neck* interpretation of the spliced target word. As predicted, participants' fixations to the referent picture *net* were delayed on hearing the spliced target word, compared with a nonmanipulated token of the target word, even when the competitor *neck* was neither present on the display nor ever mentioned in the experiment. Effects of similar splicing manipulations on the identification of word sequences have been found using a gating task, in which the spoken input is incrementally interrupted and participants have unlimited time to generate lexical hypotheses, but not with online tasks, such as the lexical-decision task (Marslen-Wilson & Warren, 1994; McQueen, Norris, & Cutler, 1999; but see Streeter & Nigro, 1979). The eye-tracking paradigm appears to provide a better temporal resolution of lexical activation over time than do other, more traditional psycholinguistic tasks. Thus, it should reveal any potential transient modulation of lexical activation.

The use of eye gaze to potential competitors as a way of evaluating their activation over time in semantically constraining contexts offers some important advantages over other measures. It does not require exposing listeners with semantically incoherent sentences, as in most brain-imaging studies on similar issues, and it does not draw listeners' attention to the semantic or phonological relationships between prime and target stimuli, as in priming methodologies. Such awareness can cause response biases whose effects are difficult to dissociate from effects attributable to lexical processes (e.g., Pitt & Shoaf, 2002; Zwitserlood, 1996).

Concerns About the Visual-World Paradigm: The Closed Set

A potentially serious limitation of the paradigm is that it requires use of a circumscribed visual world that is most often perceptually available to participants before hearing spoken input. This visual world provides the context within which the spoken input is interpreted. This closed set is certainly more constraining than the contexts provided by the psycholinguistic tasks typically used to study spoken-word recognition, which raises the concern that the closed set creates a limited set of lexical candidates that participants evaluate before, or even without, consulting more general lexical information. For example, preexposing participants to a small set of objects from which the referent of a linguistic expression will be selected may result in very specific expectations about which particular word will be mentioned, either because participants implicitly prename the pictured entities or because the pictures activate their names, making them salient in working memory. The speech input could then be exclusively analyzed with respect to these phonological expectations. Because targets are always present in the visual world, participants would never need to consider lexical candidates outside of this closed set. This strong verification strategy would be analogous to a verification set constructed from strong associates in a priming experiment in which the target was always a strong associate of the prime.

Our colleagues and we have addressed these closed-set concerns in a series of recent studies. For example, a strong form of the closed-set hypothesis predicts that frequency effects should be eliminated or strongly reduced because each of the pictured entities would be in the verification set, and each would have an equal likelihood of being the referent. However, Dahan, Magnuson, and Tanenhaus (2001) found robust effects of frequency when the display contained cohort competitors that varied in frequency (Experiment 1) and when all of the displayed pictures had unrelated names (Experiment 2). Magnuson et al. (2003) reported similar results in experiments using artificial lexicons.

Perhaps the most compelling evidence against the verification set concern comes from demonstrations of effects of neighborhood density—the number and frequency of similar-sounding words in the language—on fixations to referent pictures, even when the neighbors are not present on the display (Magnuson, 2001; Magnuson et al., 2003). These neighborhood-density effects indicate that the composition of the entire lexicon is engaged when people process the name of the referent picture. Thus, preexposure to a small set of alternatives does not appear to induce strategies that distort or neutralize critical processes known to bear on spoken-word recognition.

In addition, the hypothesis that people prename the pictured objects to constrain which picture(s) to consider while the referent's name is heard is inconsistent with the results of a recent study by Dahan and Tanenhaus (2003). On hearing the very first sounds of the name of an object, such as *snake,* participants were more likely to temporarily fixate on the picture of a competitor that shares visual features but no phonological similarity with the referent's name, such as the picture of a rope, than to fixate on a visually and phonological unrelated picture, such as a picture of a couch (see Huettig & Altmann, in press, for a similar result). It is important to note that the magnitude of this effect remained the same regardless of whether the participants had 300 ms or 1,000 ms to view the pictures before the onset of the spoken word. Fixations to a visual competitor cannot be accounted for by a match between the speech input and the name of the picture, and this casts doubt on whether participants prename the pictures at all.

In our view, preexposure to the display results in a perceptual—not lexical—analysis of the displayed objects, indexed by their spatial location. People encode the visual scene in terms of properties of the displayed objects and, most likely, what they conceptually represent. The subsequent processing of the referent's name generates lexical hypotheses, each of which is associated with conceptual and visual information. Fixations are thus directed toward the spatial locations associated with conceptually and visually matching objects.[1]

Together, these studies indicate that fixations generated to potential referent objects over time are modulated by lexical activation, mediated by the activation of the semantic or conceptual representation of words matching—even temporarily—the unfolding spoken signal. Thus, although more work is required to fully specify the linking hypothesis between observed fixations and the processing of linguistic input, the paradigm is well suited to revealing transient activation of conceptual representations of candidate words.

The use of the visual-world paradigm for investigating the interplay of context and lexical access may raise a different objection, however. Because the set of potential referents to the linguistic expression is limited to the set of visually displayed objects, one may argue that the constraints carried by a preceding verb are amplified to such extent that they do not reflect normal speech. Strong anticipations may result in abnormally reduced attention to the phonetic information from the referent's name. By itself, this is not a compelling argument, because contextual constraints can vary across normal listening situations, and the question at issue is how contextual constraints of different types combine with information from the acoustic signal. Furthermore, the constraints that context imposes in everyday speech vary to such extent that it is difficult to draw a clear distinction between normal and artificially inflated contextual constraints. Nevertheless, we will return to this objection in more detail in conjunction with the results of both Experiment 1 and Experiment 2. Crucially, we provide clear empirical evidence that the paradigm is sensitive to subtle bottom-up influences even in the face of strong contextual constraints. Thus, the paradigm should be sensitive to any short-lived bottom-up effects that would arise from context-independent access.

Overview

In the present study, we addressed the question of how verb-based thematic constraints affect lexical processing. We compared the time course of fixations to the referent and its cohort competitor in the constraining context with those in the nonconstraining context. In Experiment 1, we focused on whether there is a short time interval during which the possible referents of the target are determined solely by their bottom-up match to the input without influence from contextual constraints. In Experiment 2, we evaluated whether the system remains finely tuned to the input after contextual constraints have had a clear influence on lexical processing. We addressed this issue by introducing coarticulatory information that was biased toward the cohort competitor after the point in the speech stream at which contextual constraints had begun to affect lexical processing.

Experiment 1

Method

Participants. Twenty native speakers of Dutch, students at the University of Nijmegen, the Netherlands, participated in this experiment.

Materials. Twenty-two pairs of picturable nouns that overlapped at onset were selected. For 17 pairs, the nouns were monosyllabic and diverged after the syllable's nucleus (e.g., *bok–bot*). The remaining 5 pairs consisted of two- or three-syllable nouns sharing the entire first syllable (with the exception of *schilder–schildpad*, for which the coda of the first syllable differs). For each of the pairs, a verb was selected for which only one member of the pair would be a semantically coherent subject. We refer to the semantically coherent subject as the target and the other member of the pair as the cohort competitor. For example, the verb *klimmen* [to climb] was selected for the pair *bok–bot. Bok* [goat], the target, is a good fit as the subject of *klimmen* because a goat is a prototypical climb agent, whereas the competitor *bot* [bone] is not a plausible climb agent. In addition to the target and competitor nouns, a semantic-competitor word was selected to be a plausible subject of the verb but phonologically different from the target. For example, for the verb *klimmen*, the semantic competitor was *spin* [spider]. This semantic competitor was included to evaluate the impact of the semantic context provided by the verb on the referent's identification when phonetic information from the referent's name did not also support that alternative. Finally, a distractor word, which was semantically and phonologically different from the target word, was selected to serve as baseline (e.g., *eiland* [island]).

The thematic relationship between the verb and the target noun varied. For eight items, the target noun referred to an animate object that played the role of the agent (e.g., *bok* [goat], subject of the verb *klimmen* [to climb]). For the remaining 14 items, the target noun referred to an inanimate object that played the role of the patient (e.g., *kanon* [canon] as the subject of the verb *roesten* [to rust]). We also chose verbs for which the association with either the target or the semantic-competitor word is infrequent, as generated in word-association tasks or according to native speakers' intuitions. Word-association norms for verbs in Dutch are scarce, and data were not always available. Nevertheless, for 13 of our selected verbs, listed either as a stimulus or as an answer in van Loon-Vervoorn and van Bekkum (1991), the association frequency with the target word was 3.7% on average (ranging from 0% to 39%), and the association frequency with the semantic-competitor word was 2% on average (ranging from 0% to 14%). Thus, there was a fairly weak word association between the verb and the target or the semantic competitor. The full set of items, with English translations, is presented in Appendix A.

In addition to the 22 experimental items, 30 sets of filler items were constructed with targets that were semantically coherent subjects of the

[1] This applies primarily to displays like the ones used in the current study, which do not imply potential events. When the display and the utterances describe or imply events, the internal representation of the scene may be more richly encoded and dynamically updated (see Altmann & Kamide, in press).

verb chosen for the set. Each set consisted of four picturable nouns and a verb. The verb (e.g., *smelten* [to melt]) referred to an action of which two of the nouns (i.e., the target and its semantic competitor) could be the subject (e.g., *ijsje* [ice cream] and *boter* [butter]). The remaining two nouns were distractors. In 10 of the 30 fillers, the distractors were phonologically similar. This was done to prevent participants from developing expectations from the experimental trials that one of two phonologically similar words was likely to be the target. For the remaining 20 fillers, the distractor nouns were phonologically unrelated.

Line-drawing pictures were selected for the 120 nouns used in the experiment from various picture databases (Cycowicz, Friedman, Rothstein, & Snodgrass, 1997; Snodgrass & Vanderwart, 1980).

For each experimental item, a pair of sentences was constructed. Each sentence began with a temporal or locative adverbial or a prepositional phrase (e.g., *nog nooit* [never before], *voor het eerst* [for the first time]), which, in Dutch, constrains the conjugated verb to precede its subject. For one sentence pair, the first element was a *wh* word (*sinds wanneer* [since when]), which constrains the order of the verb and its subject in the same way. Thus, as the spoken sentence unfolded, the conjugated verb immediately followed the first constituent of the sentence and preceded its subject. The two sentences differed in that the conjugated verb was either a main verb (e.g., *klom* [climbed]) or an auxiliary or a modal verb (e.g., *is* [is], *kon* [could]; see Table 1). When the auxiliary or modal verb was conjugated, the main verb appeared at the end of the sentence (with the exception of one item, for which the omission of the conjugated verb was judged to be more natural). The main verb provided thematic and semantic constraints as to which word(s) could play the role of its grammatical subject, whereas the auxiliary or modal verb did not impose such constraints. We thus refer to these two conditions as the constraining-verb and neutral-verb conditions, respectively. The pairs of sentences used for each of the 22 experimental items and their English glosses are listed in Appendix B.

The form of the conjugated verb was also varied across the filler items. For 15 items, a main verb appeared immediately before the subject; for the remaining 15 items, an auxiliary or a modal verb preceded the subject, whereas the main verb was located near or at the end of the sentence.

All sentences were read by a male native speaker of Dutch in a soundproof room and recorded on a digital audiotape. The sentences were then digitized, edited, and labeled using the Xwaves speech-editor software (Entropic Research Laboratory, Inc.). On average, the time interval between the beginning of the sentence and the onset of the subject noun was 975 ms in the neutral-verb condition and 1,075 ms in the constraining-verb condition. The duration of the noun was virtually identical between these two conditions (283 ms and 281 ms, respectively).

Typicality norms. We collected norms to verify that the target and semantic competitors were both good fits to the thematic role assigned by the main verb. Fifteen native speakers of Dutch who did not participate in the eye-tracking experiment were asked to provide typicality ratings for the target, semantic competitor, and cohort competitor. Participants were presented with questions such as *Hoe waarschijnlijk is het voor een bok om te klimmen?* [How common is it for a goat to climb?]. A 7-point scale was used, in which 1 corresponded to *onwaarschijnlijk* [uncommon] and 7 to *waarschijnlijk* [common]. Each participant rated all three nouns associated with the same verb; the order in which each noun associated with the verb appeared in the course of the norming session was counterbalanced across participants. In addition to the 66 (22 items × 3 nouns) experimental items, 24 filler items were added (16 with an atypical subject for the verb and 8 with a more typical subject). The target received a mean rating of 5.6, the semantic competitor 5.4, and the cohort competitor 1.9. A one-way analysis of variance (ANOVA) showed a significant effect of the type of nouns: By subjects, $F_1(2, 28) = 756.2$, $p < .01$, $MSE = 0.09$; by items, $F_2(2, 42) = 190.5$, $p < .01$, $MSE = 0.52$. Pairwise comparisons showed no significant difference between the ratings of the target and its semantic competitor, $F_1(1, 14) = 3.6$, $p > .05$, $MSE = 0.04$; $F_2 < 1$. Thus, we achieved the main objectives for the materials: The cohort competitors were poor semantic fits to the role that the verb assigned its subject, and the targets and semantic competitors were good semantic fits.

Table 1
Stimulus Example From Experiment 1

Visual stimuli	
Target: bok [goat] Cohort competitor: bot [bone] Semantic competitor: spin [spider] Distractor: eiland [island]	
Condition	Auditory stimuli
Constraining verb	Nog nooit klom een bok zo hoog [Never before climbed a goat so high]
Neutral verb	Nog nooit is een bok zo hoog geklommen [Never before has a goat climbed so high]

Note. English glosses appear in brackets.

Procedure and design. Participants were first familiarized with the 120 pictures to ensure that they identified them as intended. Each picture appeared on a computer screen, along with its printed name. Participants were instructed to familiarize themselves with each picture and to press a response button to proceed to the next picture. After this part of the experiment, the eye-tracking system was set up.

Participants were seated at a comfortable distance from the computer screen. The eye-tracking system was mounted and calibrated. Eye movements were monitored with an SMI Eyelink (SensoMotoric Instruments, GmbH) eye-tracking system, sampling at 250 Hz. Spoken sentences were presented to the participants through headphones. The structure of a trial was as follows. First, a central fixation point appeared on the screen for 500 ms, followed by a blank screen for 600 ms. Then, a 5 × 5 grid with four pictures and four geometric shapes appeared on the screen as the auditory presentation of a sentence was initiated. Participants used the computer mouse to move the object mentioned in the spoken sentence above or below its adjacent geometric shape. The positions of the pictures were randomized across four fixed positions of the grid, whereas the geometric shapes appeared in fixed positions on every trial. Participants' fixations for the entire trial were completely unconstrained, and participants were under no time pressure to perform the action. Once the picture had been moved, the experimenter pressed a button to initiate the next trial. Every five trials, a central fixation point appeared on the screen, allowing for an automatic drift correction of the calibration.

Two lists were created by varying which of the two sentences (i.e., the neutral-verb or constraining-verb sentence) was presented for each of the 22 experimental items. Within each list, 11 experimental items were assigned to each condition. For each list, three random orders were created, with the constraint that five of the filler trials were presented at the beginning of the experiment to familiarize participants with the task and procedure. Participants were randomly assigned to each list, with an approximately equal number of participants assigned to each of the within-list orders.

Coding procedure. The data from each participant's right eye were analyzed and coded for fixations, saccades, and blinks. (For one participant, data for the left eye were used because of calibration problems with the right eye.) Onsets and offsets of saccades were determined using the thresholds for motion (0.2 deg), velocity (30 deg of visual angle/s), and acceleration (8,000 deg/s^2). Fixation duration corresponded to the time interval between two successive saccades, and fixation position was determined by averaging the *x* and *y* coordinates of the eye position recorded during the fixation. Graphical analysis software performed the mapping between the position of fixations, the mouse movements, and the pictures present on each trial and displayed them simultaneously. Each fixation was represented by a dot associated with a number denoting the order in which the fixation had been produced; the onset and duration of each fixation were available for each fixation dot. The timing of fixations was established relative to the onset of the target word (i.e., the name of the referent) in the spoken utterance. Fixations were coded as directed to the target picture, to one of the competitor pictures, to the distractor picture, or to anywhere else on the screen. Fixations that fell within the cell of the grid in which a picture was presented were coded as fixations to that picture. On each experimental trial, fixations were coded from the onset of the target word until participants had clicked on the target picture with the mouse, which was taken to reflect the participants' identification of the referent. Fixations performed subsequently (often to the geometric shapes) were not coded. In most cases, participants were fixating the target picture when clicking on it. In the rare cases in which participants did not look at the target picture while clicking on it, an earlier long fixation to the target picture was taken as indicating the referent's identification, and the coding of the trial ended with that fixation.

Results and Discussion

The data from 11 trials (2.5% of the data) were excluded from the analyses because of track loss (1 trial) or because participants moved the wrong picture without correcting their choice (10 trials, 6 in the neutral-verb condition and 4 in the constraining-verb condition). The proportion of fixations to each picture or location (i.e., target picture, competitor pictures, distractor picture, or elsewhere) over time (in 10-ms time intervals) for each condition and each participant was calculated by adding the number of trials in which a picture type was fixated during a 10-ms time interval and dividing it by the total number of trials in which a fixation to any picture or location was observed during this time interval (thus excluding in this count the trials in which a blink or a saccade occurred during that 10-ms time interval). Proportion values for each picture and each time interval were averaged across participants, separately for each condition.

Figure 1 presents the proportions of fixations to the target, the cohort competitor, the semantic competitor, and the distractor over time, for the neutral-verb condition (Figure 1A) and the constraining-verb condition (Figure 1B). In the neutral-verb condition, fixations to the target and cohort competitor increased with a similar slope from 200 ms until around 350 ms after target-word onset. Fixations to the target then continued to rise while fixations to the cohort competitor began to drop. The onset of the rise in competitor fixations is consistent with the well-established delay in planning and launching a saccadic eye movement in a display with multiple potential targets, estimated to average 200 ms. The dotted lines indicate the mean time interval over which fixations reflect the processing of the target word (i.e., from 200 ms after target-word onset until the average target-word offset with the added 200-ms delay). As is apparent on the graph, the rise and fall in cohort-competitor fixations occurred while the target word was heard and processed, providing initially consistent and subsequently inconsistent information with the cohort competitor. This pattern is similar to that observed in previous eye-tracking studies using cohort competitors (e.g., Allopenna et al., 1998; Dahan, Magnuson, & Tanenhaus, 2001).

The pattern of fixation proportions was strikingly different in the constraining-verb condition. Fixations to the target and cohort competitor diverged very early. Moreover, fixations to the cohort competitor were almost indistinguishable from those to the distractor. Finally, fixations to the semantic competitor did not differ much from those to the distractor (and in fact showed no difference between the neutral-verb and constraining-verb conditions).

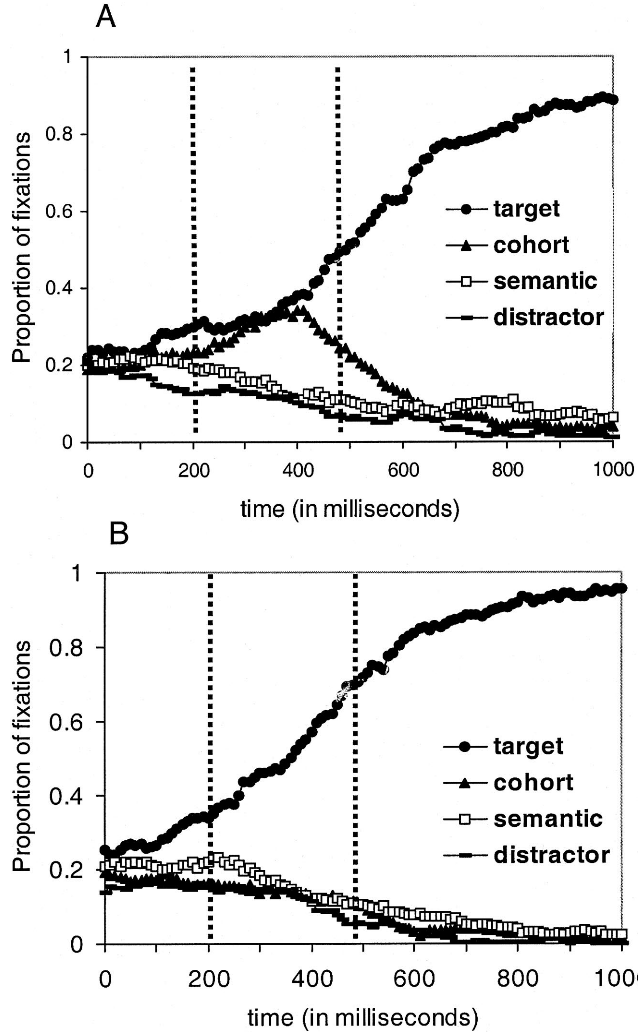

Figure 1. Proportion of fixations to the target, the cohort competitor, the semantic competitor, and the distractor over time from the onset of the target word (in milliseconds) in the neutral-verb condition (A) and in the constraining-verb condition (B). The dotted lines indicate the mean time interval over which fixations reflect the processing of the target word (with a 200-ms delay for saccadic latency).

In order to statistically test this pattern, fixation proportions were averaged over a particular time window for each subject and item and were submitted to a one-way repeated-measures ANOVA by subject (F_1) or by item (F_2). The time interval chosen for the primary analysis extended from 200 ms to 500 ms after target-word onset. This time window begins 200 ms after the onset of the target word because this is the earliest point at which fixations driven by information from the target word are expected. The time window extends over 300 ms, which roughly corresponds to the mean duration of the target word (283 ms for the neutral-verb condition and 281 ms for the constraining-verb condition). This dependent measure represents, for a particular picture, the probability of being fixated during a particular time interval.

The mean proportion of fixations to the target was higher in the constraining-verb condition (51.5%) than it was in the neutral-verb condition (36.2%), $F_1(1, 19) = 24.4$, $p < .01$; $F_2(1, 21) = 15.3$,

$p < .01$. Conversely, the proportion of fixations to the cohort competitor was higher in the neutral-verb condition (28.5%) than it was in the constraining verb condition (13.6%), $F_1(1, 19) = 31.3$, $p < .01$; $F_2(1, 21) = 11.0$, $p < .01$. The proportion of fixations to the semantic competitor did not differ significantly between the two verb conditions (14.6% in the neutral-verb condition and 16% in the constraining-verb condition, $F_1 < 1$; $F_2 < 1$). Finally, in order to directly compare the cohort-competitor fixations with the distractor fixations in each of the two verb conditions, we conducted a two-way ANOVA (Picture × Condition). This analysis revealed a main effect of picture, $F_1(1, 19) = 19.1$, $p < .01$; $F_2(1, 21) = 6.4$, $p < .05$; of condition, $F_1(1, 19) = 20.9$, $p < .01$; $F_2(1, 21) = 9.1$, $p < .01$; and most crucially, a significant interaction, $F_1(1, 19) = 15.4$, $p < .01$; $F_2(1, 21) = 7.6$, $p < .05$. The proportion of fixations to the cohort competitor was greater than the proportion of fixations to the unrelated distractor in the neutral-verb condition (28.5% vs. 10.9%), $F_1(1, 19) = 32.9$, $p < .01$; $F_2(1, 21) = 9.6$, $p < .01$, but not in the constraining-verb condition (13.6% vs. 11.8%), $F_1 < 1$; $F_2 < 1$.

In order to establish that the effect of the verb context was observed from the earliest moments of lexical processing, we conducted an additional analysis over a time interval extending from 200 ms to 300 ms after target-word onset. The mean proportion of fixations to the target was higher in the constraining-verb condition than it was in the neutral-verb condition, although the effect was marginal by items, $F_1(1, 19) = 8.33$, $p < .01$; $F_2(1, 21) = 3.5$, $p = .075$. The proportion of fixations to the cohort competitor was significantly higher in the neutral-verb than it was in the constraining-verb condition, both by subjects and by items, $F_1(1, 19) = 14.42$, $p < .01$; $F_2(1, 21) = 4.96$, $p < .05$. Finally, the proportion of fixations to the semantic competitor did not differ between the two verb conditions, $F_1 < 1$; $F_2 < 1$.

Fixation proportions, as defined here, represent the probability of fixating a picture over a defined time interval. This probability is modulated by new fixations or by fixations that began earlier and continue during this time window. Other measures have been used that rely on a link between linguistic processing and shifts of attention toward a particular object. Altmann and colleagues (Altmann & Kamide, 1999; Kamide et al., 2003) analyzed the proportion of trials in which a saccade to a particular object is initiated while a linguistic expression is heard. We conducted similar analyses on the present data, computing the proportion of trials in which a saccade to the cohort competitor was initiated while the target word was processed. This time interval was item specific and extended from 200 ms after target-word onset to 200 ms after target-word offset. On average, such saccades were initiated on 24.3% of the trials in the neutral-verb condition and on only 10.6% of the trials in the constraining-verb condition, $F_1(1, 19) = 9.9$, $p < .01$; $F_2(1, 21) = 9.8$, $p < .01$. This effect described in terms of saccadic eye movements is consistent with what was revealed by the fixation-proportion analysis.

To summarize, the results show an immediate effect of verb context. When the target word (i.e., the subject of the verb) was preceded by a semantically constraining main verb, the cohort competitor was not considered for recognition more than was a distractor picture, despite the phonetic overlap between its name and the acoustic information. This finding suggests that constraints provided by the context influenced lexical processing from the earliest moments of lexical access, with no observable time interval during which the meaning of a contextually incongruent competitor was activated. Furthermore, during the critical time window over which the referent's name was processed, the semantic competitor was not considered more after a constraining verb than after a neutral verb, indicating that the identification of the referent was based on phonetic and semantic information, both supporting the target.

As is apparent in Figure 1, however, fixations to the target picture in the constraining-verb condition began to increase about 100 ms after the onset of the target word, which is surprisingly early given standard assumptions about the programming time for eye movements and given results obtained in previous studies using the same paradigm. This pattern suggests that these early eye movements might have been anticipatory eye movements generated primarily on the basis of information from the verb. In fact, given that the duration of the spoken word(s) separating the offset of the verb and the onset of the subject noun phrase was 97 ms on average, the timing of the target increase indicates that most of these fixations were programmed immediately after verb offset. Unexpectedly, fixations to the target but not to the semantic competitor were affected by such verb-based anticipations. Recall that the typicality norms showed only a small nonsignificant tendency for the targets to be judged to be more typical subjects than the semantic competitors. It is possible, however, that in the constraining-verb contexts, the portion of the sentence up through the verb might have made the target more predictable than the semantic competitor. If that were the case, the predictability difference for some items might have been sufficient to result in more anticipatory eye movements to the target picture. In order to evaluate this possibility, we conducted a second norming study.

Sixteen participants saw the four pictures associated with each of the 22 experimental trials and a written transcription of the constraining-verb sentence truncated just before the subject noun. Their task was to decide which picture was likely to be mentioned next. The order with which the 22 trials were presented, as well as the position of each picture on the screen, was varied across four lists, with an approximately equal number of participants tested in each list. Participants chose the target picture 56.3% of the time, the semantic competitor 38.1%, the cohort competitor 3.1%, and the distractor 2.6%. Both the target and semantic competitor were selected significantly more often than either the cohort competitor or the distractor, which did not differ from each other. The difference between the target and semantic-competitor responses was statistically significant (with a two-tailed binomial test, $z = 3.46$, $p < .05$). Thus, the initial part of the sentence, when combined with the verb, introduced an overall bias in favor of the target. This bias differed across items, however. Out of the 22 items, 12 showed a strong bias in favor of the target (chosen 80.2% of the time), 7 showed a bias in favor of the semantic competitor (chosen 76.8% of the time), and the remaining 3 items were roughly equibiased between the target and the semantic competitor.

In order to test whether the target bias might explain the relatively large proportion of early looks to the target in the constraining-verb condition, we correlated, over the 22 items, the proportion of looks to the target picture in the 0–200-ms time interval and the proportion of target choices in the norming study. There was a strong positive correlation, $r(20) = .83$, $p < .01$, indicating that many of these fixations came from anticipatory eye movements driven by the verb. However, the predictability bias had a short-lived effect on target fixations: When computed over the 200–500-ms time interval, the correlation between the target-fixation proportion and the percentage of times the target picture was selected in the norms was small and nonsignificant ($r = .14$). Nevertheless, the presence of an early bias in favor of the target raises the possibility that participants mostly used predictions from the verb to establish the referent of the verb's subject without processing the spoken information on the target word. To address this issue, we separately analyzed fixations from the 12 items that were target biased (as established by the predictability norms) and the remaining 10 nontarget-biased items. Figure 2 presents the proportion of fixations over time to the target, the cohort competitor, the semantic competitor, and the distractor, in the constraining-verb condition, for each of these item sets. For the target-biased items (Figure 2A), the proportion of fixations to the target began to rise sharply in the first 200 ms after target onset and continued throughout the target word. In contrast, for the nontarget-biased items (Figure 2B) the proportion of fixations to the target remained roughly at the same level as the cohort competitor until about 200 ms, whereas the proportion of fixations to the semantic competitor rose notably. At 200 ms, which is precisely when the uptake of phonetic information from the target word is expected to be revealed in fixations, given the saccade-latency estimate, the tendency reversed, and the proportion of fixations to the target rose sharply, whereas that to the semantic competitor began to drop. Crucially, there was little, if any, delay in fixations to the target for the nontarget-biased items compared with the target-biased items. In fact, the fixation curves reached a comparable level by 300 ms and were indistinguishable by 350 ms. In both item sets, the fixation proportions to the cohort competitor remained low and in fact were equivalent to those to the unrelated distractor. Thus, the impact of phonetic information from the target word itself was clearly observable on those trials in which anticipatory eye movements initially favored the semantic competitor.

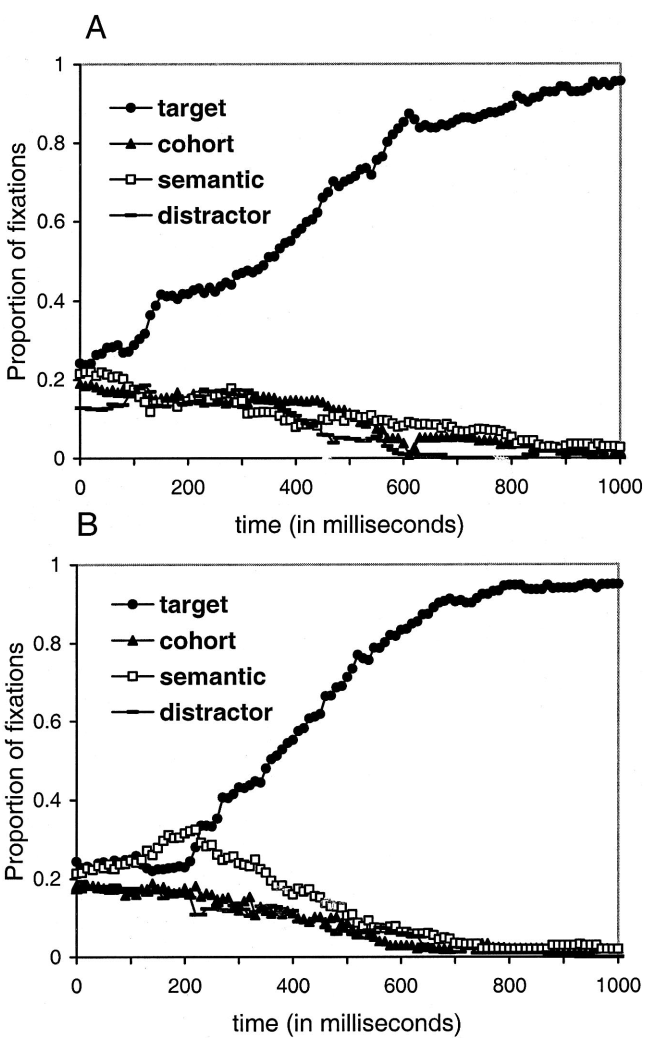

Figure 2. Proportion of fixations to the target, the cohort competitor, the semantic competitor, and the distractor over time from the onset of the target word (in milliseconds) in the constraining-verb condition, for the target-biased items (A) and the nontarget-biased items (B).

We conducted one-way ANOVAs on fixation proportions to the target and cohort pictures over the 200–500-ms time interval for the 10 nontarget-biased items and replicated the pattern of results found for the full item set, with only marginally reliable verb effects by items on target fixations: $F_1(1, 19) = 5.86$, $p < .05$; $F_2(1, 9) = 4.14$, $p = .07$; and on cohort-competitor fixations: $F_1(1, 19) = 11.18$, $p < .01$; $F_2(1, 9) = 3.57$, $p = .09$.

These results demonstrate that only fixations taking place in the first 200 ms were anticipatory responses based on the early part of the sentence and the verb. Such a verb-based anticipatory effect is consistent with Altmann and Kamide's (1999) finding. From 200 ms on, fixations reflect the integration of the phonetic information from the target word and the semantic context established by the verb.

The nontarget-biased items provide strong evidence against a strategy-based explanation for the effects of the verb-based constraint. For these items, participants were most likely to be fixating on the unrelated semantic competitor during the first 200 ms and were roughly equally likely to be fixating on either the target or the cohort competitor. As participants processed the initial portion of the target word, fixations to the cohort did not increase relative to the unrelated distractor, demonstrating that the input was immediately integrated with the verb-based contextual constraint even when the target was not the most expected entity.

The results from Experiment 1 strongly support the hypothesis that the activation of lexical representations during the recognition of a spoken word is affected by immediate semantic integration with context. In the neutral context, participants were equally likely to fixate on the cohort competitor and the target. This result indicates that, as the word unfolded, semantic information associated with the target and cohort competitor was equally activated. However, when the preceding verb introduced thematic constraints, participants were no more likely to look at a contextually incongruent cohort competitor than at an unrelated distractor. Thus, there was no evidence that initial activation of lexically based semantic information was determined by purely form-driven fit with the phonetic input. Instead, fixations were determined by the combination of form-based fit with the unfolding input and consistency with the verb-based thematic constraints. The results of Experiment 1 demonstrate immediate integration of contextual and signal-driven constraints, which is consistent with continuous mapping but not with the access–selection hypothesis.[2] However, the results do not address the equally important question of whether contextual constraints affect sensitivity to subsequent input. In Experiment 2, we focused on this question.

Experiment 2

In the access–selection view, part of the rationale for having an initial candidate set that is established solely on the basis of the goodness of fit with the phonetic input is allowing words to be recognized even when they are embedded in a context that makes them implausible. However, if one assumes a *continuous* integration of multiple sources of information, such as contextual and phonetic sources, there is no need for delayed influence of contextual information. When contextual biases are weak, the weighting of potential lexical candidates will primarily reflect goodness of fit to the phonetic input. However, when contextual constraints are strong, a contextually consistent candidate could immediately benefit from its fit to context. Subsequent strong phonetic information could still change the balance of evidence in favor of an alternative candidate. Many parallel constraint-based models of syntactic processing adopt this approach, immediately integrating multiple sources of constraint but maintaining sensitivity to multiple alternatives (Jurafsky, 1996; MacDonald, Pearlmutter, & Seidenberg, 1994; McRae, Spivey-Knowlton, & Tanenhaus, 1998).

In order to ascertain whether integration of multiple cues is continuous, we evaluated the processing of phonetic input that arrives after the integration with context has already begun. Numerous studies have demonstrated that listeners use coarticulatory information in vowels to favor some lexical candidates over others, for example, favoring *bok* over *bot* even before the burst of the final stop consonant *k* is heard (e.g., Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Marslen-Wilson & Warren, 1994; McQueen et al., 1999; Warren & Marslen-Wilson, 1987; Whalen, 1984, 1991). We altered the acoustic realization of the target word by introducing coarticulatory information in the vowel that was biased toward the cohort competitor. This information was introduced by using cross-splicing, combining, for example, the initial consonant and vowel from *bot* ([bone], the name of the cohort competitor) with the burst of the final consonant from *bok* ([goat], the name of the target). Because of coarticulation, the final portion of the vowel /ɔ/ contains phonetic cues that anticipate the articulation of the following consonant. We predicted that these cues would temporarily bias the interpretation toward a candidate (e.g., *bot*) that was later disfavored by the unambiguous final portion of the word (e.g., the released final consonant *k*). This cross-splicing manipulation tested the sensitivity to phonetic information occurring after semantic constraints began to bias interpretation without exposing participants to semantically odd sentences.

The effects of the cross-splicing manipulation were tested in both the neutral-verb and the constraining-verb conditions. In the neutral-verb condition, the target and cohort competitor were each consistent with the context. Consequently, as the target word unfolded and the coarticulatory information in the vowel was heard, we predicted that there would, temporarily, be more fixations to the competitor picture with the altered version of the target word (e.g., *bo*[*t*]*k*) than with the original one, replicating the results of Dahan, Magnuson, Tanenhaus, and Hogan (2001). We predicted that this effect would emerge at about the same time as when coarticulatory information started favoring the target interpretation in the unaltered version of the target word.

The critical condition was the constraining-verb condition. Following the results of Experiment 1, we expected an initial bias toward the target picture early in the target word, with the proportion of looks to the cohort competitor being similar to the proportion of looks to the unrelated distractor. Of primary interest were the effects of processing the coarticulatory information in the vowel. If the system remains tuned to the signal even after the context has favored a possible referent, there should be more fixations to the cohort competitor when the realization of the target word is altered—hence providing phonetic cues in support of the cohort competitor—compared with when the target word is intact. Moreover, the effects of misleading information in the constraining-verb condition should not be delayed compared with the effects of misleading coarticulatory information in the neutral-verb condition, reflecting a continuous integration of all sources of information. Such an outcome would demonstrate that the processor remains closely tuned to the acoustic input even when enough contextual and phonetic information has accumulated to allow identification of the referent. Influential mispronunciation-detection studies revealed poorer performances when the mispronunciation occurred late in a word compared with early in a word (e.g., Cole, Jakimik, & Cooper, 1978; Marslen-Wilson & Welsh, 1978). This result was interpreted as evidence for reduced analysis of the phonetic input once enough information to identify a word is available. However, this issue is worth reexamining in light of the criticisms that have been raised about the mispronunciation-detection task (see van Donselaar, 1996, for a review).

Method

Participants. Seventy-two native speakers of Dutch, students at the University of Nijmegen, participated in the experiment. None of the participants had taken part in Experiment 1 or in the norming experiments.

Materials. We eliminated 5 of the 22 items used in Experiment 1 because the target and cohort-competitor names could not be easily cross-spliced. We excluded items for which the target or the cohort competitor was embedded in its counterpart (e.g., *koe* being embedded in *koets*). In addition, the pair *paraplu–paprika* was excluded because the first syllable of *paprika* receives a primary stress, whereas that of *paraplu* receives a secondary stress. The mismatch generated by splicing the first syllable of one word into the final portion of the other for this item would be of a different nature than the mismatch generated by the splicing performed on the other items. Finally, the pair *schilder–schildpad* was also excluded because the first syllables of the two items did not fully overlap. The 17 remaining items, as well as all the filler items, were identical to those used in Experiment 1, with the exception of the auditory stimuli associated with each item.

A new recording was made using the same speaker as in Experiment 1. For each experimental item, the speaker read four different sentences: two neutral-verb sentences with the target word or the cohort competitor as the subject (e.g., "Nog nooit is een bok/bot zo hoog geklommen" [Never before has a goat/bone climbed so high]) and two constraining-verb sentences with the target word or the cohort competitor as the subject (e.g., "Nog nooit klom

[2] Proponents of the access–selection view could argue that fixations to displayed pictures in the task used here reflect the outcome of the selection process after the context has operated on an initial form-based candidate set. However, the timing of the fixations would imply no temporal delay between access and selection, making the theoretical distinction between these two processes difficult to motivate.

een bok/bot zo hoog" [Never before climbed a goat/bone so high]). For 14 of the 30 fillers, only one sentence mentioning the target word was read, with an auxiliary or a modal (7 fillers) or a main verb (7 fillers) preceding the subject noun. For each of the remaining 16 fillers, two sentences were produced, one mentioning the target word and the other a competitor word (e.g., "Van bovenaf keek de hagedis/hazewind verbaasd naar de nieuwe bezoeker" [From above looked the lizard/greyhound surprisingly at the new visitor]). For 9 of these 16 fillers, a main verb preceded the subject noun; for the remaining 7, an auxiliary or a modal verb preceded the subject.

All sentences were recorded on a Digital/Audio tape. They were then digitized, edited, and labeled using the Xwaves speech-editor software (Entropic Research Laboratory, Inc.). For each verb-condition sentence, two spliced versions were created by splicing the initial part of the sentence mentioning either the target or the cohort competitor up to and including the vowel of the first syllable onto the final part of another token of the sentence mentioning the target word (see Table 2 for an illustration of this procedure). This process yielded two versions of lexically identical sentences: one version in which the presplice portion of the subject noun originated from the same word as did the postsplice portion, the target word (hereafter, the identity-spliced version), and another version in which the presplice portion of the subject noun originated from a different word than did the postsplice portion, the cohort competitor (hereafter, the cross-spliced version).

Splicing points were first established through visual inspection of the waveforms and auditory feedback, locating the offset of the vowel in the first syllable. (For two items, namely *band–bank* and *wolk–wolf,* the splicing point was located after the first consonant of the coda cluster.) Some of the splicing points established this way had to be adjusted because the splicing between segments from different sentences or tokens of the same sentence resulted in acoustic artifacts, such as clicks or other obvious distortions. Of the 102 splicing points (17 Experimental Items × 6 Splicing Points), 33 required adjustment. These adjustments were roughly evenly distributed across verb conditions. The size of the adjustment was quite small (14.8 ms on average, varying from 0.04 ms to 33.3 ms). The portion of the sentence preceding the target-word onset was, on average, 988 ms in the neutral-verb condition (987 ms in the identity-spliced version and 988 ms in the cross-spliced version) and 1,141 ms in the constraining-verb condition (1,136 ms in the identity-spliced version and 1,146 ms in the cross-spliced version). The duration of the target word was, on average, 322 ms in the neutral-verb condition (318 ms in the identity-spliced version and 326 ms in the cross-spliced version) and 311 ms in the constraining-verb condition (310 ms in the identity-spliced version and 312 ms in the cross-spliced version). The duration of the presplice portion was 186 ms on average, with only small differences across the four subconditions.

A similar splicing manipulation was used with the fillers. Identity-spliced versions were created by splicing two tokens of the same sentence for the 14 fillers for which only a sentence, mentioning the target word, had been produced. Cross-spliced versions were created for the 16 fillers for which a version mentioning the target word and a version mentioning a competitor word had been produced.

Design. Experiment 2 contained four subconditions, Verb (neutral vs. constraining) × Splicing (identity vs. cross). However, the number of experimental items was quite small (17). In order to maximize the number of items tested in each condition per participant, we adopted a design in which the critical factor, Splicing, was a within-subjects factor, whereas Verb was a between-subjects factor. As a result, each participant was tested on 17 experimental trials, all of them in either the neutral-verb condition or the constraining-verb condition; 9 of them were in the identity-splicing condition, whereas the other 8 were in the cross-splicing condition. Four lists were constructed by varying the subcondition in which each item was tested. Fillers were identical for all four lists, with 16 in the constraining-verb condition and 14 in the neutral-verb condition. Three random orders per list were created, varying the position of each of the pictures on the grid and the trial orders. Participants were randomly assigned to each list; an approximately equal number of participants were assigned to each random order. As a consequence of this design, the number of subjects tested per condition per item was equivalent between Experiments 1 and 2, despite the fact that more subjects were tested in Experiment 2.

The testing and coding procedures were identical to those in Experiment 1.

Results and Discussion

The data from 35 trials (2.9% of the data) were excluded from the analyses because participants moved the wrong picture without correcting their choice. It was not surprising that this occurred most often for the cross-spliced version in the neutral-verb condition (23 trials), although such errors were also found in other subconditions (5 trials in the identity-spliced version of the neutral-verb condition and 7 trials in the cross-spliced version of the constraining-verb condition). The proportion of fixations to each picture or location (i.e., target picture, competitor pictures, distractor picture, or elsewhere) over time (in 10-ms time intervals) for each condition and each participant was calculated in the same way as in Experiment 1.

Table 2
Illustration of the Stimuli Construction for Experiment 2

Sentence type and condition	Stimuli
Constraining-verb condition	
Original	Nog nooit klom een bok zo hoog$_a$
	Nog nooit klom een bot zo hoog$_b$
	Nog nooit klom een bok zo hoog$_c$
Spliced	
Identity-splicing condition	Nog nooit klom een bo$_a$k zo hoog$_c$
Cross-splicing condition	Nog nooit klom een bo$_b$k zo hoog$_c$
Neutral-verb condition	
Original	Nog nooit is een bok zo hoog geklommen$_d$
	Nog nooit is een bot zo hoog geklommen$_e$
	Nog nooit is een bok zo hoog geklommen$_f$
Spliced	
Identity-splicing condition	Nog nooit is een bo$_d$k zo hoog geklommen$_f$
Cross-splicing condition	Nog nooit is een bo$_e$k zo hoog geklommen$_f$

Note. The subscripts in the spliced sentences indicate the original sentences from which each portion of the spliced sentences originated. Target = bok; cohort competitor = bot.

Figure 3 presents the proportions of fixations over time to the target picture and its cohort competitor. In the neutral-verb condition (Figure 3A), there is a clear effect of the cross-splicing manipulation. The fixation proportion to the cohort competitor and the target both started rising around 200 ms, but the fixation proportion to the cohort competitor reached a higher level and remained higher in the cross-spliced version than in the identity-spliced version. Conversely, the fixation proportions to the target picture rose more slowly in the cross-spliced version than in the identity-spliced version. This pattern indicates that the interpretation of the target word was strongly influenced by coarticulatory information, temporarily favoring the cohort competitor over the target in the cross-spliced version. Furthermore, the target and cohort-competitor curves in the identity-spliced versions (e.g., when the target word is *bo[k]k*), indicated with filled symbols on Figure 3, started diverging around 350 ms after target onset. Fixation proportions to the target continued to rise, whereas those to the cohort competitor began to drop, indicating that participants' interpretation of the target word's referent favored the target over its competitor. This is thus the point at which the fixation proportions reflect the uptake of coarticulatory information in the target word's vowel, which started favoring the target over its cohort competitor. If one factors in a 200-ms delay in programming and launching an eye movement, this estimate is closely time locked to the final part of the presplice portion of the target word, which, averaged across the stimuli, had a duration of 186 ms.

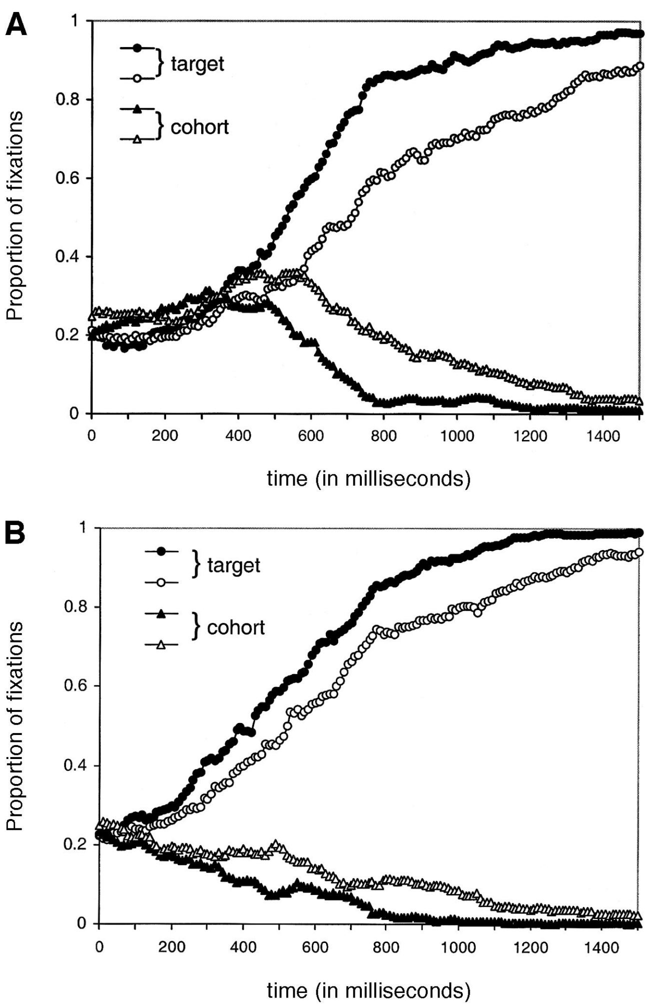

Figure 3. Proportion of fixations to the target and the cohort competitor over time from the onset of the target word (in milliseconds) in the identity-spliced condition (filled symbols) and the cross-spliced condition (empty symbols), in the neutral-verb context (A) and in the constraining-verb context (B).

In the constraining-verb condition (Figure 3B), the fixation proportion to the target picture started rising in the first 200 ms, while the fixation proportion to the cohort competitor started dropping. This pattern is consistent with what we observed in Experiment 1, in which the presence of a constraining verb before the subject noun appeared to exclude the semantically inconsistent cohort competitor despite its initial phonetic overlap with the target word. However, from about 350 ms on, the slope with which the fixation proportion to the cohort competitor dropped was slower in the cross-spliced version than it was in the identity-spliced version. Thus, participants tended to fixate on the cohort competitor more often and for longer when the target word's vowel contained coarticulatory information favoring this cohort competitor than when it did not. Crucially, this pattern was observed despite the fact that the preceding constraining verb had rendered the cohort competitor a poor potential referent for the target word. The fixation proportion to the target picture showed a complementary pattern, indicating that the recognition of the target word was delayed by the presence of short-lived mismatching coarticulatory information, even though the preceding verb context and the initial sounds of the target word had established it as a good referent for the target word. This finding suggests that the interpretation of the target word reflects a continuous integration of contextual and phonetic constraints.

These effects were evaluated statistically using a two-way (Verb $\times$ Splicing) ANOVA on the fixation proportions to the cohort competitor and to the target separately, over the 350–500-ms time interval. This interval extends roughly from the offset of the presplice portion (plus a 200-ms delay) to the offset of the target word (plus a 200-ms delay). During this interval, fixations resulting from processing the coarticulatory information should be observed. Although the effect clearly extends over a larger time window, as shown in Figure 3, finding an effect of cross-splicing during this narrow window most convincingly demonstrates the tight time locking between the processing of phonetic information and its consequences on the referential interpretation of the target word.

The fixation proportion to the cohort competitor was higher in the cross-spliced version than it was in the identity-spliced version in the neutral-verb condition (34.0% vs. 27.7%) as well as in the constraining-verb condition (18.3% vs. 9.8%). There was a main effect of verb $F_1(1, 70) = 69.2$, $p < .01$; $F_2(1, 16) = 40.6$, $p < .01$; a main effect of splicing, $F_1(1, 70) = 14.6$, $p < .01$; $F_2(1, 16) = 5.2$, $p < .05$; and importantly, no interaction, $F_1 < 1$; $F_2 < 1$, indicating a statistically equivalent effect of the splicing manipulation whether the verb preceding the target word was semantically constraining or not.

Analyses on the fixation proportions to the target picture revealed a comparable pattern. Fixation proportions were lower in the cross-spliced version than they were in the identity version in the neutral-verb condition (29.4% vs. 37.0%) as well as in the constraining-verb condition (41.5% vs. 51.7%). There was a main effect of verb, $F_1(1, 70) = 19.8$, $p < .01$; $F_2(1, 16) = 18.2$, $p < .01$; a main effect of splicing, $F_1(1, 70) = 13.6$, $p < .01$; $F_2(1, 16) = 5.5$, $p < .05$; and no interaction, $F_1 < 1$; $F_2 < 1$.

Finally, fixations to the semantic competitor while the target word was being processed did not vary as a function of the verb, replicating the results from Experiment 1. A two-way ANOVA (Verb × Splicing) on fixation proportions to the semantic competitor over a time window extending from 200 ms to 500 ms after target onset revealed no main effect of verb or splicing and no interaction (all Fs < 1).

We conducted additional analyses to support the claim that splicing had an immediate effect on fixations. Fixation proportions to target and cohort competitor were computed, by subjects and by items, over the 200–300-ms time interval following the item-specific splicing point and were submitted to two-way (Verb × Splice) ANOVAs. They revealed a pattern similar to what was found on the 350–500-ms window, that is, a significant main effect of verb and splicing, with no interaction between these factors. The splicing effect on cohort-competitor fixations was only marginally significant by items, $F_2(1, 16) = 3.3$, $p = .09$.

We also performed analyses on the proportion of trials in which saccades to the cohort-competitor picture were initiated during the item-specific postsplicing time interval, adjusted by the 200-ms saccade-programming delay (i.e., from the splicing point + 200 ms to the target-word offset + 200 ms). Saccadic eye movements to the cohort competitor were launched more often in the cross-spliced condition than they were in the identity-spliced condition 12.5% vs. 7.2%), $F_1(1, 70) = 10.4$, $p < .01$; $F_2(1, 16) = 7.6$, $p < .05$; and more often in the neutral-verb condition than they were in the constraining-verb condition (12.3% vs. 7.2%), $F_1(1, 70) = 6.8$, $p < .05$; $F_2(1, 16) = 4.2$, $p = .05$, with no interaction between these two factors.

Thus, the results demonstrate that the system remained finely tuned to the bottom-up input, responding to small short-lived coarticulatory information that was inconsistent with the preferred referent and consistent with a contextually incongruent lexical candidate. The system was equally sensitive to the relatively subtle constraints provided by coarticulatory cues when one lexical candidate was strongly biased, as in the constraining-verb condition, and when both candidates were equally consistent with the context, as in the neutral-verb condition. This result demonstrates that the early effect of semantic constraints on lexical interpretation does not result in reduced or delayed analysis of the phonetic input. It also reveals that short-lived coarticulatory information can temporarily change the balance of evidence in favor of an alternative that is inconsistent with the verb's thematic constraints. This result is important because it shows that late-arriving phonetic information is a powerful source of constraint that can overcome initial biases that are based on the preceding phonetic input and semantic constraints.

This demonstration of sensitivity to the input after an initial contextual bias also validates the use of the visual-world paradigm for studying constraint integration. As discussed earlier, the paradigm requires the use of a circumscribed set of potential referents that is visually available to participants before hearing spoken input. One legitimate concern is that the influence of sentential context on lexical processing is abnormally amplified when the semantic restrictions that this context may convey operate on such a small set. In other words, by hearing the verb *climb* in the context of a goat, a spider, a bone, and an island, participants build abnormally strong expectations about the referent's identity. If such expectations are set, the argument goes, it is of no surprise that referents that do not match these expectations are never considered. Experiment 2's results showed that a nonexpected referent, such as the cohort competitor, can be reconsidered in the course of hearing the referent's name. This is remarkable because the coarticulatory information in the cross-spliced stimuli was only weakly inconsistent with the target, and it arrived after the input had clearly distinguished between the target and the semantic competitor, which were the only contextually congruent targets. Moreover, the misleading coarticulatory information was quickly overridden by clear phonemic information that was consistent with the target. Finally, participants never encountered trials in which the ultimate target was incongruent with the context. This demonstrates that the concurrence of a small set of visually available referents and a constraining verb did not set expectations that could not be revised. Thus, finding a cross-splicing effect in the constraining-verb condition provides strong evidence that participants' referential interpretations of the target word resulted from a continuous integration of phonetic and semantic constraints, rather than from task-specific guessing strategies based on selectively attending to the predicted referents.

General Discussion

The current experiments make two primary contributions to our understanding of lexical processing in spoken utterances. First, they establish that semantic mapping and integration are continuous. Signal-driven fixations to pictures compatible with a target word began approximately 200 ms after its onset. In neutral contexts, participants were equally likely to fixate a target and its cohort competitor. However, when the preceding verb context established thematic constraints, fixations were limited to the referent that matched both the thematic and the phonetic constraints. From the earliest moments of lexical processing, then, listeners accessed lexical meanings and integrated them with the relevant context.

It is important to note that these results are not incompatible with the results of other studies in which contextual constraints appear to have more delayed effects. In contrast to the contexts that are typically used in studies of lexical-ambiguity resolution, such as those of Huettig and Altmann (in press) and Zwitserlood (1989), our contexts established verb-based thematic constraints that become available as soon as the verb is processed and place well-defined constraints on a limited set of potential arguments.

Second, the current studies demonstrate that the processing system remains continuously sensitive to the input. In neutral-verb contexts, looks to the referent of the target word began to diverge from those to its cohort competitor around 350 ms after target-word onset, approximately 200 ms after the onset of coarticulatory information in the vowel. This was also the point in time at which looks to the cohort competitor briefly rose above looks to the target when the stimuli were cross-spliced to introduce temporarily misleading coarticulatory information that was more consistent with the cohort than with the target. In constraining-verb contexts, looks to the cohort in the cross-spliced version of the target word began to diverge from those in the identity-spliced version at around the same point in time, namely 350 ms. An important finding was that both the timing and the magnitude of the splicing effect were similar in both the neutral-verb and constraining-verb contexts. Thus, the system remains sensitive to phonetic details of the input, even after the input and context have strongly converged on a single lexical candidate.

The present study makes an important methodological contribution as well. In the visual world paradigm, as in any psycholinguistic experiment, participants adopt some strategy to perform the required task, and one can ask whether this distorts how the processes of interest are performed. The visual display, in concurrence with a linguistic context, may set expectations that are abnormally strong, and this would result in an artificially reduced sensitivity to the phonetic information on the referent's name. Experiment 2's results revealed subtle modulations in participants' interpretation of the phonetic input, even when this input favored a referent that was not supported by prior expectations. Thus, the visual world paradigm should be a useful tool for studying how multiple sources of constraints are accessed and integrated as words are processed in spoken utterances.

The results of this study present a serious challenge to the access–selection framework. Our evidence that contextual constraints can affect even the earliest moments of mapping the input onto the developing representation is more consistent with the hypothesis that spoken-language processing proceeds by continuously integrating the signal with other relevant constraints. When contextual constraints are weak, the initial mapping will be primarily driven by bottom-up phonetic constraints. However, when contextual constraints are strong, they will have immediate effects.

The continuous-mapping perspective abandons an assumption that has been central to the access–selection framework, namely the idea that the fundamental goal of lexical processing is to identify the single best-fitting lexical candidate, in other words, to recognize the words in the speech stream. If the goal of lexical processing is word recognition, the access–selection approach is appealing in that it seems to provide for an optimally efficient system. Lexical candidates that are consistent with the input make available semantic and syntactic information. Selection can then proceed using a mix of top-down and bottom-up constraints. As evidence converges, a single lexical candidate wins, and the recognition process is completed, often on the basis of 200 ms or less of the input word.

However, we would argue that the goal of lexical processing is not to make a decision about which words are present in the input stream. Rather, the goal is to make lexically specific information available for ongoing computations about *interpretation*. Interpretation implies the mapping of linguistic expressions onto objects or events in the conceptual or real world. This mapping cannot be achieved by merely attributing a predefined meaning to a sound form. As argued by Johnson-Laird (1987), among others, all open-class words have an indeterminate meaning that can only restrict a semantic framework for the referential interpretation of the word; the interpretation is further specified on the basis of contextually appropriate inferences. For example, the particular referential interpretation of the unambiguous word *fish* in the sentence "A swimmer was attacked by a fish" is equivalent to that of *shark.* Under this view, both the context and the sound form of the current word provide information that can constrain referential computation or interpretation of the word *fish.* Their respective involvement in lexical processing is thus viewed as simultaneous rather than as functionally and temporarily distinct.

What classes of models are consistent with continuous mapping? Bayesian models of information integration provide one possible approach. In Bayesian models, multiple sources of information are evaluated simultaneously and in a probabilistic manner to achieve the optimal interpretation of the signal. Furthermore, the propagation of constraints takes place continuously, so that the total support for each alternative is continuously updated. This view is embodied in models of automatic speech recognition that adopt a Bayesian framework (see, e.g., Rabiner & Juang, 1993) as well as in Massaro's fuzzy logical model of perception (e.g., Massaro, 1987, 1994, 1996). Distributed models that focus on computation and integration of multiple sources of information are a second possible approach (cf. McClelland, 1996). According to this view, the distinction between linguistic levels, often envisioned as separate processing modules, is blurred by assuming a pattern of activation over an ensemble of units to encompass several levels of representations simultaneously. The spoken-word recognition model proposed by Gaskell and Marslen-Wilson (1997, 1999, 2002), in which a single set of nodes encodes both the form and the meaning of the currently processed word, represents a step in this direction.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38,* 419–439.

Altmann, G. T. M. (1999). Thematic role assignment in context. *Journal of Memory and Language, 41,* 124–145.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73,* 247–264.

Altmann, G. T. M., & Kamide, Y. (in press). Now you see it, now you don't: Mediating the mapping between language and the visual world. In J. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action.* New York: Psychology Press.

Boland, J. E. (1997). The relationship between syntactic and semantic processes in sentence comprehension. *Language and Cognitive Processes, 12,* 423–484.

Boland, J. E. (2001). Verb argument structure rapidly guides visual attention while listening. *Abstracts of the 42nd Annual Meeting of the Psychonomic Society, 6,* 114.

Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. N. (1995). Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language, 34,* 774–806.

Cole, R. A., Jakimik, J., & Cooper, W. E. (1978). Perceptibility of phonetic features in fluent speech. *Journal of the Acoustical Society of America, 64,* 44–56.

Connine, C. M. (1987). Constraints on interactive processes in auditory word recognition: The role of sentence context. *Journal of Memory and Language, 26,* 527–538.

Connine, C. M., Blasko, D. G., & 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.

Connine, C. M., Blasko, D. G., & Wang, J. (1994). Vertical similarity in spoken word recognition: Multiple lexical activation, individual differences, and the role of sentence context. *Perception and Psychophysics, 56,* 624–636.

Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal words of spoken sentences. *Journal of Cognitive Neuroscience, 6,* 256–266.

Cycowicz, Y. M., Friedman, D., Rothstein, M., & Snodgrass, J. G. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology, 65,* 171–237.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology, 42,* 317–367.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes, 16,* 507–534.

Dahan, D., & Tanenhaus, M. K. (2003). *Activation of visually-based conceptual representations during spoken-word recognition.* Manuscript submitted for publication.

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language, 27,* 429–446.

Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language, 44,* 516–547.

Fischer, B. (1992). Saccadic reaction time: Implications for reading, dyslexia and visual cognition. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 31–45). New York: Springer-Verlag.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language, 37,* 58–93.

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes, 12,* 613–656.

Gaskell, M. G., & Marslen-Wilson, W. D. (1999). Ambiguity, competition, and blending in spoken word recognition. *Cognitive Science, 23,* 439–462.

Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Representation and competition in the perception of spoken words. *Cognitive Psychology, 45,* 220–266.

Hallet, P. E. (1986). Eye movements. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and human performance* (pp. 10-1–10-112). New York: Wiley.

Huettig, F., & Altmann, G. T. M. (in press). Language-mediated eye movements and the resolution of lexical ambiguity. In M. Carreiras & C. Clifton (Eds.), *The on-line study of sentence comprehension: Eye-tracking, ERP, and beyond.* New York: Psychology Press.

Johnson-Laird, P. N. (1987). The mental representation of the meaning of words. *Cognition, 25,* 189–211.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science, 20,* 137–194.

Kako, E., & Trueswell, J. C. (2000). Verb meaning, object affordance, and the incremental restriction of reference. *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 256–261). Hillsdale, NJ: Erlbaum.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence for anticipatory eye movements. *Journal of Memory and Language, 49,* 133–156.

Lucas, M. (1999). Context effects in lexical access: A meta-analysis. *Memory & Cognition, 27,* 385–398.

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

Magnuson, J. S. (2001). *The microstructure of spoken word recognition.* Unpublished doctoral dissertation, University of Rochester.

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General, 132,* 202–227.

Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science, 189,* 226–227.

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

Marslen-Wilson, W. D. (1989). Access and integration: Projecting sounds onto meaning. In W. D. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 3–24). Cambridge, MA: MIT Press.

Marslen-Wilson, W. (1993). Issues of process and representation in lexical access. In G. T. M. Altmann & R. Shillcock (Eds.), *Cognition models of speech processing: The second Sperlonga meeting* (pp. 187–210). Hove, England: Erlbaum.

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

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10,* 29–63.

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry.* Hillsdale, NJ: Erlbaum.

Massaro, D. W. (1994). Psychological aspects of speech perception. Implications for research and theory. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 219–263). San Diego, CA: Academic Press.

Massaro, D. W. (1996). Integration of multiple sources of information in language processing. In T. Inui & J. L. McClelland (Eds.), *Attention and performance XVI. Information integration in perception and communication* (pp. 397–432). Cambridge, MA: MIT Press.

McClelland, J. L. (1996). Integration of information: Reflections on the theme of attention and performance XVI. In T. Inui & J. L. McClelland (Eds.), *Attention and performance XVI. Information integration in perception and communication* (pp. 633–656). Cambridge, MA: MIT Press.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86,* B33–B42.

McQueen, J. M., Norris, D., & Cutler, A. (1999). Lexical influence in phonetic decision making: Evidence for subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1363–1389.

McRae, K., Ferretti, T. R., & Amyote, L. (1997). Thematic roles as verb-specific concepts. *Language and Cognitive Processes, 12,* 137–176.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language, 38,* 283–312.

Miller, J. L., & Eimas, P. D. (1995). Speech perception: From signal to word. *Annual Review of Psychology, 46,* 467–492.

Pitt, M. A., & Shoaf, L. (2002). Revisiting bias effects in word-initial phonological priming. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 1120–1130.

Pollatsek, A., Rayner, K., & Collins, W. E. (1984). Integrating pictorial information across eye movements. *Journal of Experimental Psychology: General, 113,* 426–442.

Rabiner, L., & Juang, B.-H. (1993). *Fundamentals of speech recognition.* Englewood Cliffs, NJ: Prentice Hall.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422.

Rayner, K., Slowiaczek, M. L., Clifton, C., Jr., & Bertera, J. H. (1983). Latency of sequential eye movements: Implications for reading. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 912–922.

Saslow, M. G. (1967). Latency for saccadic eye movement. *Journal of the Optical Society of America, 57,* 1030–1033.

Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowski, M. (1982). Automatic access of the meaning of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology, 14,* 489–537.

Simpson, G. B. (1994). Context and the processing of ambiguous words. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 359–374). San Diego, CA: Academic Press.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 174–215.

Streeter, L. A., & Nigro, G. N. (1979). The role of medial consonant transitions in word perception. *Journal of the Acoustical Society of America, 65,* 1533–1541.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18,* 645–659.

Tabossi, P. (1988). Effects of context on the immediate interpretation of unambiguous nouns. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 153–162.

Tabossi, P., Colombo, L., & Job, R. (1987). Accessing lexical ambiguity: Effects of context and dominance. *Psychological Research, 49,* 161–167.

Tabossi, P., & Zardon, F. (1993). Processing ambiguous words in context. *Journal of Memory and Language, 32,* 359–372.

Tanenhaus, M. K., Garnsey, S. M., & Boland, J. (1990). Combinatory lexical information and language comprehension. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 383–408). Cambridge, MA: MIT Press.

Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence of multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior, 18,* 427–440.

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research, 29,* 557–580.

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

Trueswell, J. C., & Kim, A. E. (1998). How to prune a garden path by nipping it in the bud: Fast priming of verb argument structure. *Journal of Memory and Language, 39,* 102–123.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 528–553.

van Berkum, J. J. A., Zwitserlood, P., Hagoort, P., & Brown, C. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research, 17,* 701–718.

van den Brink, D., Brown, C., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience, 13,* 967–985.

van Donselaar, W. (1996). Mispronunciation detection. *Language and Cognitive Processes, 11,* 621–628.

van Loon-Vervoorn, W. A., & van Bekkum, I. J. (1991). *Woordassociatie lexicon.* Amsterdam: Swets & Zeitlinger.

Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 394–417.

Warren, P., & Marslen-Wilson, W. (1987). Continuous uptake of acoustic cues in spoken word recognition. *Perception and Psychophysics, 41,* 262–275.

Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception and Psychophysics, 35,* 49–64.

Whalen, D. H. (1991). Subcategorical phonetic mismatches and lexical access. *Perception and Psychophysics, 50,* 351–360.

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

Zwitserlood, P. (1996). Form priming. *Language and Cognitive Processes, 11,* 589–596.

Appendix A

Experimental Items

Target	Cohort competitor	Semantic Competitor	Distractor	Verb
*baby [baby]	beker [beaker]	worm [worm]	harp [worm]	kruipen [to crawl]
*bok [goat]	bot [bone]	spin [spider]	eiland [island]	klimmen [to climb]
*boom [tree]	boot [boat]	kind [child]	kwast [paintbrush]	groeien [to grow]
clown [clown]	klauw [claw]	papegaai [parrot]	fiets [bicycle]	roepen [to shout]
*haan [rooster]	haak [hook]	ridder [knight]	muts [woolen hat]	vechten [to fight]
*kabouter [gnome]	kado [present]	hond [dog]	pan [frying pan]	luisteren [to listen]
*pad [toad]	pak [suit]	leeuw [lion]	zon [sun]	springen [to jump]
*schaap [sheep]	schaats [skate]	giraf [giraffe]	snorkel [snorkel]	bijten [to bite]
schilder [painter]	schildpad [turtle]	wasmachine [washing machine]	kers [cherry]	werken [to work]
*band [tire]	bank [couch]	fles [bottle]	muur [wall]	rollen [to roll]
*bel [bell]	berg [mountain]	telefoon [telephone]	laars [boot]	rinkelen [to jingle]
*dak [roof]	das [tie]	emmer [bucket]	web [web]	lekken [to leak]
*hemd [undershirt]	helm [helmet]	gordijn [curtains]	liniaal [ruler]	scheuren [to rip]
*kaas [cheese]	kaars [candle]	framboos [raspberry]	sleutel [key]	beschimmelen [to get moldy]
*kanon [cannon]	kameel [camel]	slot [lock]	mijter [mitre]	roesten [to rust]
*kast [wardrobe]	kat [cat]	envelop [envelope]	vlag [flag]	bevatten [to contain]
koets [carriage]	koe [cow]	grasmaaier [lawn mower]	schelp [shell]	rijden [to ride]
lamp [lamp]	lam [lamb]	spiegel [mirror]	draak [dragon]	breken [to break]
paraplu [umbrella]	paprika [bell pepper]	krant [newspaper]	kam [comb]	vouwen [to fold]
*pijp [pipe]	pijl [arrow]	kachel [heater]	mossel [mussel]	roken [to smoke]
*pot [jar]	pop [doll]	gieter [sieve]	rits [zipper]	bevatten [to contain]
*wolk [cloud]	wolf [wolf]	veer [feather]	piano [piano]	zweven [to float]

Note. Asterisks indicate the items used in Experiment 2. English glosses appear in brackets.

Appendix B

Constraining-Verb and Neutral-Verb Sentences

Vandaag kruipt de **baby** een stuk verder [Today crawls the **baby** a bit further]
Vandaag is de **baby** een stuk verder gekropen [Today has the **baby** crawled a bit further]
Nog nooit klom een **bok** zo hoog [Never before climbed a **goat** so high]
Nog nooit is een **bok** zo hoog geklommen [Never before has a **goat** climbed so high]
Sinds wanneer groeit een **boom** zo snel? [Since when grows a **tree** so fast?]
Sinds wanneer kan een **boom** zo snel groeien? [Since when can a **tree** grow so fast?]
Zoals altijd roept de **clown** naar het publiek [As always shouts the **clown** at the audience]
Zoals altijd ging de **clown** naar het publiek toe roepen [As always was the **clown** shouting at the audience]
Deze keer vecht de **haan** zijn laatste wedstrijd [This time fights the **rooster** his last game]
Deze keer heeft de **haan** zijn laatste wedstrijd gevochten [This time has the **rooster** fought his last game]
Voor het eerst luistert een **kabouter** naar het geluid [For the first time listens a **gnome** to the sound]
Voor het eerst zal een **kabouter** naar het geluid luisteren [For the first time will a **gnome** listen to the sound]
Over het algemeen springt een **pad** niet zo hoog [Usually jumps a **toad** not so high]
Over het algemeen zal een **pad** niet zo hoog springen [Usually will a **toad** jump not so high]
Vrijwel altijd bijt er een **schaap** in het voer [Almost always bites a **sheep** the food]
Vrijwel altijd heeft er een **schaap** in het voer gebeten [Almost always has a **sheep** bitten the food]
Al drie uur achtereen werkte de **schilder** in zijn atelier [For three hours on end worked the **painter** in his studio]
Al drie uur achtereen heeft de **schilder** in zijn atelier gewerkt [For three hours on end has the **painter** worked in his studio]
Vaak rolt zo'n **band** nog een heel stuk verder [Often rolls such a **tire** considerably further]
Vaak zal zo'n **band** nog een heel stuk verder rollen [Often will such a **tire** roll considerably further]
Tegen het midden van de dag rinkelt de **bel** heel vaak [Around noon jingles the **bell** very often]
Tegen het midden van de dag heeft de **bel** heel vaak gerinkeld [Around noon has the **bell** jingled very often]
Tegenwoordig lekt een **dak** niet meer [Nowadays leaks a **roof** no more]
Tegenwoordig hoeft een **dak** niet meer te lekken [Nowadays needs a **roof** not to leak anymore]
Onder te sterke spanning scheurt een **hemd** gemakkelijk [Under too much strain rips an **undershirt** easily]
Onder te sterke spanning zal een **hemd** gemakkelijk scheuren [Under too much strain will an **undershirt** rip easily]
In een bedompte ruimte beschimmelt een **kaas** veel sneller [In a humid room gets moldy a **cheese** a lot faster]
In een bedompte ruimte is een **kaas** veel sneller beschimmeld [In a humid room will a **cheese** get moldy a lot faster]
Vol in de open lucht roest zo'n **kanon** behoorlijk [Out in the open air rusts such a **cannon** considerably]
Vol in de open lucht gaat zo'n **kanon** behoorlijk roesten [Out in the open air is going such a **cannon** to rust considerably]
Op dit moment bevat de **kast** wat oude spullen [At this moment contains the **wardrobe** some old stuff]
Op dit moment zou de **kast** wat oude spullen moeten bevatten [At this moment should the **wardrobe** contain some old stuff]
Elke dag reed de **koets** door het land [Everyday rode the **carriage** across the country]
Elke dag heeft de **koets** door het land gereden [Everyday has the **carriage** ridden across the country]
Plotseling brak er een **lamp** in de straat [Suddenly broke a **lamp** in the street]
Plotseling is er een **lamp** in de straat gebroken [Suddenly has a **lamp** been broken in the street]
Hierna vouwde de **paraplu** beter [After this folded the **umbrella** better]
Hierna kon de **paraplu** beter vouwen [After this could the **umbrella** fold better]
Op dat tijdstip rookte de **pijp** al lang niet meer [At that point in time had stopped smoking the **pipe** a long time ago]
Op dat tijdstip moet de **pijp** al lang niet meer hebben gerookt [At that point in time should the **pipe** have stopped smoking a long time ago]
Gisteren bevatte de **pot** alleen wat meel [Yesterday contained the **jar** only some flour]
Gisteren heeft de **pot** alleen wat meel bevat [Yesterday has the **jar** contained only some flour]
Enkele minuten geleden zweefde de **wolk** boven ons [Some minutes ago floated the **cloud** above us]
Enkele minuten geleden was de **wolk** boven ons [Some minutes ago was the **cloud** above us]

Note. The conjugated verb is underlined; the target word (playing the role of the subject) is in bold. English glosses appear in brackets.

Received May 2, 2003
Revision received September 12, 2003
Accepted October 3, 2003 ■