Journal of Experimental Psychology:
Human Perception and Performance
2007, Vol. 33, No. 4, 960–977

0096-1523/07/$12.00 DOI: 10.1037/0096-1523.33.4.960

Effects of Syntactic Expectations on Speech Segmentation

Sven L. Mattys, James F. Melhorn, and Laurence White
University of Bristol

Although the effect of acoustic cues on speech segmentation has been extensively investigated, the role of higher order information (e.g., syntax) has received less attention. Here, the authors examined whether syntactic expectations based on subject–verb agreement have an effect on segmentation and whether they do so despite conflicting acoustic cues. Although participants detected target words faster in phrases containing adequate acoustic cues ("spins" in *take spins* and "pins" in *takes pins*), this acoustic effect was suppressed when the phrases were appended to a plural context (*those women take spins/*takes pins* [with the asterisk indicating a syntactically unacceptable parse]). The syntactically congruent target ("spins") was detected faster regardless of the acoustics. However, a singular context (*that woman *take spins/takes pins*) had no effect on segmentation, and the results resembled those of the neutral phrases. Subsequent experiments showed that the discrepancy was due to the relative time course of syntactic expectations and acoustics cues. Taken together, the data suggest that syntactic knowledge can facilitate segmentation but that its effect is substantially attenuated if conflicting acoustic cues are encountered before full realization of the syntactic constraint.

Keywords: speech segmentation, allophony, syntactic expectations, lexical competition

How listeners handle the relative continuity of the spoken input has been the focus of a great deal of attention. Finding word boundaries in connected speech is thought to involve both signal-driven and knowledge-driven processes. Signal-driven processes include the use of sublexical cues probabilistically associated with word boundaries, such as allophonic and coarticulatory word-onset variants (e.g., Davis, Marslen-Wilson, & Gaskell, 2002; Mattys, 2004; Quené, 1992, 1993; Salverda, Dahan, & McQueen, 2003), phonotactic regularities (e.g., McQueen, 1998; Vitevitch & Luce, 1999), and stress placement (e.g., Cutler & Butterfield, 1992; Cutler & Norris, 1988). Reliance on these cues has been documented both in isolation and in combination (for a review, see Mattys, White, & Melhorn, 2005). Knowledge-driven processes, too, have been shown to facilitate segmentation. However, the empirical evidence has been mostly confined to the lexical level, whereby word boundaries are identified once competition among partly overlapping candidates settles on a lexically optimal parsing solution (e.g., Gaskell & Marslen-Wilson, 1997; McClelland & Elman, 1986; Norris, 1994).

Higher order knowledge (e.g., semantics, syntax) has been investigated as well, but mostly insofar as it allows listeners to anticipate or influence lexical selection (e.g., Altmann & Kamide, 1999; Dahan & Tanenhaus, 2004) or disambiguate lexically ambiguous candidates (e.g., Seidenberg, Tanenhaus, Leiman, & Bienkowsky, 1982). How higher order knowledge contributes to speech *segmentation*, however, has been difficult to establish (Sanders & Neville, 2000, 2003). The available evidence for the effects of higher order knowledge on lexical segmentation–activation strongly favors a continuous and interactive view of contextual–lexical integration, in which contextual expectations have an online supplementary or inhibitory effect on the activation of lexical candidates, independent of the fit with sensory information (cf. above references; Mattys, Pleydell-Pearce, Melhorn, & Whitecross, 2005; but, for task-specific effects, see Borsky, Shapiro, & Tuller, 2000). When it comes to specifying the unique contribution of syntax, investigation is often impeded by substantial confounding of syntax and semantics, as syntactic constraints often imply semantic restrictions (e.g., verbs usually refer to actions) and syntactic violations almost always result in a loss of meaning.[1]

Sven L. Mattys, James F. Melhorn, and Laurence White, Department of Experimental Psychology, University of Bristol, Bristol, United Kingdom.

This research was supported by Biotechnology and Biological Sciences Research Council Grant 7/S18783 to Sven L. Mattys.

Correspondence concerning this article should be addressed to Sven L. Mattys, Department of Experimental Psychology, University of Bristol, 12A Priory Road, Bristol BS8 1TU, United Kingdom. E-mail: sven.mattys@bris.ac.uk

[1] Researchers have sometimes attempted to tease apart syntax and semantics by comparing syntactically acceptable but semantically meaningless sentences (e.g., "He prepared at the back hand to pair up his robbers") with either syntactically and semantically acceptable sentences (e.g., "She played the drum in a rock and roll band") or random word strings (e.g., "Be place prefer the was city it and sure be perfume"; Van Petten & Kutas, 1991; see also Marslen-Wilson & Tyler, 1980; Tyler & Marslen-Wilson, 1986; Tyler & Warren, 1987; Tyler & Wessels, 1983). Although this method is informative for general comparisons among lexical, syntactic, and semantic influences on speech processing, it has limitations. First, the syntax–semantics subtractive method rests on the assumption that the difference between syntactically acceptable and unacceptable meaningless sentences is uniquely accounted for by syntax. However, incongruent semantics could be more taxing in the former than the latter, because acceptable syntax is likely to encourage a more thorough search for meaning than is a random string of words. Thus, the results could reflect not only syntactic effects but also additional semantic processing. Second, because the subtractive method usually involves indiscriminate syntactic violation, it does not allow the locus of the syntactic effects or the mechanism that causes them to be specified. In particular, it cannot discriminate syntactic effects spanning several words or phrases from those involving local dependencies. This shortcoming can be minimized if distinctions between conditions are limited to local ambiguities (e.g., [with the underlined portion under investigation and + and − denoting the presence or absence, respectively, of semantic or syntactic information], Sem+/Syn+: "In order to recycle <u>bo</u>ttles . . . "; Sem−/Syn+: "In order to lefatal <u>bo</u>kklers . . ."; Sem−/Syn−: "Ah ilgen di lefatal <u>bo</u>kkerth . . ."; Sanders & Neville, 2000, 2003).

Such limitations can be attenuated when the effects of local syntactic violations are estimated from recognition performance on specific target words. For instance, word monitoring is found to be substantially slower when the speech material preceding a target word (in italics) violates an expected prepositional dependency (e.g., "The children were hoping for *snow*" vs. *"The children were hoping *snow*" [with the asterisk indicating a syntactically unacceptable parse]; Baum, 1989, 1991) or subject–verb number agreement (e.g., "The women carry the child and eat *ice cream*" vs. *"The women carry the child and eats *ice cream*" [examples translated from Dutch]; Haarmann & Kolk, 1994). Although studies such as the ones cited above provide evidence for online sensitivity to syntactic violations during speech comprehension, they do not specify whether or how syntax assists the speech processor for segmentation purposes.

The goal of the present study was to investigate whether syntactic knowledge affects the perception of word boundaries and, if so, how it combines with other word-boundary cues. We recently showed that lexical and contextual knowledge tends to override sublexical cues in case of segmentation conflict (Mattys, White, & Melhorn, 2005). For example, in the sentence "The two players left in the tournament will contest the final," artificially removing the coarticulatory information between "con" and "test"—a manipulation shown to provide a strong word-boundary cue (Johnson & Jusczyk, 2001; Mattys, 2004)—did not facilitate the segmentation of the lexically and contextually inconsistent word "test." However, decoarticulation effects were apparent when the stimuli were played in a background of mild noise. On the basis of these and other results, we concluded that, in intelligible speech conditions, listeners give greater weight to lexical and contextual information than to sublexical cues such as coarticulation, phonotactics, and stress. Although syntax was loosely included in the lexical–contextual tier, its effect on segmentation was not directly investigated. The present experiments represent an attempt to do so and to position syntactic expectations relative to acoustic information in the hierarchical structure of segmentation cues.

Using the subject–verb third-person number agreement rule in English, we explored the conditions under which a pivotal /s/ is treated as a verb inflection as opposed to the first phoneme of the subsequent object, given orthogonal manipulations of syntax and acoustics. The critical stimuli consisted of pairs of near-homophonous verb–object phrases such as "take spins" and "takes pins," which were either presented as such (neutral condition) or preceded by a singular or plural subject phrase (for an illustration, see Table 1). Thus, in the neutral condition, the only cues to word junctures were acoustic (e.g., segmental lengthening, aspiration, and allophonic cues). In the singular and plural contexts, however, syntactic expectations were introduced in the form of a subject phrase varying in number, with the pivotal /s/ consistent with a verb inflection in the singular case (e.g., "That woman takes pins") or with the initial segment of the object in the plural case (e.g., "Those women take spins").

Of primary interest was how listeners handle conflicts between acoustic and syntactic information ("That woman take#spins" and "Those women takes#pins" [with # indicating an acoustic juncture]). An account of segmentation that gives greater weight to syntax than to sublexical cues (e.g., Mattys, White, & Melhorn, 2005) predicts that such conflicts should be solved syntactically rather than acoustically. Alternative patterns of results, however, would suggest a more graded approach to cue integration and a refinement of the hierarchy. We tested the dominance hypothesis using word monitoring, a technique known to be sensitive to both signal- and knowledge-specific information and, of particular interest, to syntactic complexity and violations (for an overview, see Kilborn & Moss, 1996). Participants monitored the presence of a prespecified target word (e.g., "spins" and, on a separate trial, "pins") in the utterances described above (i.e., the neutral, singular, and plural syntactic contexts for "take#spins" vs. "takes#pins"). Monitoring speed was taken as an indication of ease of word segmentation. The acoustic cues were created by concatenating test stimuli recorded in isolation, (e.g., "take" and "spins" or "takes" and "pins"; see the *Design and procedure* section of Experiment 1; for earlier uses of this technique, see also Mattys, 2004; Mattys, White, & Melhorn, 2005). An advantage of the concatenation technique is that it affords greater control over the actual location of the intended boundaries (e.g., Johnson & Jusczyk, 2001), ensuring that the acoustic cues thought to be associated with word boundaries are realized to a perceivable extent. However, because the decoarticulation–concatenation technique might artificially enhance the salience of the acoustic information, the experiment also included a natural (*natural-allophony*) condition. The recording distinction (concatenation vs. natural allophony) was entered in the analyses mostly as a control factor.

Table 1
Examples of Utterances in Experiments 1 and 2

Utterance	Consistent target: Acoustics	Consistent target: Syntax
Experiment 1		
Neutral		
take#spins	spins	
takes#pins	pins	
Singular		
*That woman take#spins	spins	pins
That woman takes#pins	pins	pins
Plural		
Those women take#spins	spins	spins
*Those women takes#pins	pins	spins
Experiment 2 (syntax supplemented with semantics)		
Neutral		
take#spins	spins	
takes#pins	pins	
Singular		
*The new machine at the bowling alley take#spins	spins	pins
The new machine at the bowling alley takes#pins	pins	pins
Plural		
Boys like it when their go-karts take#spins	spins	spins
*Boys like it when their go-karts takes#pins	pins	spins

Note. Utterance–target consistency is indicated in the two rightmost columns. Asterisks denote cases in which reliance on acoustic cues would lead to incorrect subject–verb agreement. # = an acoustic juncture.

Experiment 1

Method

Participants. Ninety native speakers of British English, undergraduate or graduate students at the University of Bristol (Bristol, United Kingdom), received course credit or a small honorarium for their participation in the experiment. None reported a history of speech or hearing difficulties. They were randomly assigned to either the concatenation ($n = 45$) or the natural-allophony ($n = 45$) condition.

Materials. We selected 15 verb–object phrases containing an ambiguous pivotal /s/, in which the /s/ could be treated as either the initial segment of the noun (e.g., "take#spins") or the singular inflection of the verb (e.g., "takes#pins"). All but 2 verbs were monosyllabic, and all nouns were monosyllabic. The onset of the non-/s/-initial noun was a plosive in 10 cases (e.g., "pace"), a nasal in 2 cases (e.g., "miles"), and a liquid in 2 cases (e.g., "lips"). In 1 case, the syllable was onsetless ("eels"). The average log CELEX frequencies of the /s/-initial nouns (e.g., "spins") and of the non-/s/-initial nouns (e.g., "pins") were 2.49 and 2.12, respectively, $t(14) = 1.63$, $p = .13$ (Baayen, Piepenbrock, & Gulikers 1995). Similarly, there was no significant difference between the phonotactic probability of the diphone straddling one word boundary (e.g., "take spins") and that of the diphone straddling the alternative word boundary (e.g., "takes pins"): Average numbers of occurrences were 2,641 versus 2,186 (anywhere in a word), $t(14) = .51$, $p = .61$, and 299 versus 537 (at syllable boundaries), $t(14) = -1.61$, $p = .13$ (from CELEX). The two acoustic versions of each phrase ("take#spins" and "takes#pins") were created following the procedure described in the *Design and procedure* section below.

The verb–object pairs of phrases were placed in three syntactic contexts. The neutral condition featured the phrases in isolation. In the singular and plural conditions, the phrases were preceded by a singular subject (e.g., "That woman") or a plural subject (e.g., "Those women"), respectively. A set of stimuli is shown in Table 1. The subject phrases were chosen to be as semantically neutral as possible relative to the verb–object phrases. To minimize the possibility that the verb (e.g., "take") would be interpreted as a noun modifying the object rather than a verb, the subject phrases all included determiners or quantifiers (e.g., "the," "one," "many"). Although this did not entirely preclude a noun–noun interpretation of the subject–verb group (MacDonald, 1993), the prosodic contour of the utterances helped convey the intended syntactic categories (Millotte, René, Wales, & Christophe, 2007). All test materials are listed in Appendix A.

The test utterances were presented with the /s/-initial and non-/s/-initial words as targets to monitor (e.g., "spins" and "pins"). In addition, 60 target-present filler utterances included 18 two-word phrases and 42 whole sentences. The two-word phrases contained the target in the initial position, with half of them leading to a plausible syntactic parse (e.g., "heals" in "heals flesh") and the other half leading to an illegitimate parse (e.g., "ditch" in "ditches vans"). Of the 42 whole sentences, half contained the target in an initial portion (e.g., "job" in "His job is to mail post") and the other half contained the target in a medial portion (e.g., "have" in "The grocers have plums").

One hundred and twenty target-absent trials consisted of a mixture of short phrases and whole sentences paired with a monosyllabic target word not present in the utterance. Finally, 10 practice trials made of original materials were assembled as a training sample of the actual test conditions.

Design and procedure. The verb–object pairs were recorded separately from the subject phrases. The acoustic cues between the verb and the object were created either through concatenation or through natural allophony. In the concatenation condition, the speaker was asked to produce the verb–object phrase at the end of a padding sentence (e.g., "He said takes pins," "He said take spins") with a mouth closure and a short silent pause between the verb and the object (for a similar procedure, see Mattys, 2004; Mattys, White, & Melhorn, 2005). The pause and the padding portion of the sentence were then edited out and the concatenated verb–object phrase saved (e.g., "take#spins," "takes#pins"). In the natural-allophony condition, we followed a similar procedure, except that the speaker was asked to produce the verb–object phrase in a natural fashion, without pausing. Thus, here, word-boundary cues resulted from natural allophonic variants of the segments around the junctures. Quantitative and qualitative analyses of the test utterances are reported in Appendix B.

To create the singular and plural utterances, we concatenated the singular and plural subject phrases, recorded separately, to the verb–object phrases. The juncture between the subject phrase and the verb–object phrase, though decoarticulated, sounded like well-enunciated natural speech. To prevent the acoustic makeup of the test utterances from being used as an indication of a target-present trial, we also produced and edited the fillers and target-absent trials following the above procedure, with alternate positions for the decoarticulation–concatenation points.

All stimuli were recorded in a sound-attenuated booth by a male native speaker of standard southern British English. They were digitized (16 bit A/D) at 32 kHz. On output, the utterances were converted to analog form (16 bit D/A, 32 kHz) and delivered over good-quality headphones. To minimize the number of repetitions of a given verb–object phrase, we assigned participants to the three context conditions (neutral, singular, and plural) in a Latin square design rotating across stimulus sets. Thus, although each participant contributed data to all three conditions, he or she never heard a given set in more than one condition. The position of the acoustic juncture (e.g., "take#spins" vs. "takes#pins") and the target to detect (e.g., "spins" vs. "pins") were orthogonal within-subject variables.

Each participant completed 240 trials: 60 test trials (15 sets $\times$ 2 acoustic conditions $\times$ 2 target conditions), 60 target-present filler trials, and 120 target-absent filler trials. Trials were pseudorandomized, with at least 5 intervening trials between trials from the same set. Participants were randomly assigned to the concatenation condition or the natural-allophony condition and tested individually in a quiet room. They were seated in front of a computer monitor and wore headphones. They were told that, on each trial, a letter string would be presented on the computer monitor in front of them, followed by an utterance played over the headphones. They were instructed to push a button as soon as they heard the target in the utterance and to push another button if the utterance did not contain the target. They were told not to pay attention to grammar and pronunciation in the detection task. Both speed and accuracy were emphasized.

On each trial, a visual target appeared for 1 s on the computer monitor. Then, the utterance was played, with the target remaining visible until the end of the utterance. After the participant pressed a response key, or 3 s after the end of the utterance, there was a 1-s wait before the next visual target was displayed.

Results

In an attempt to maximize comparability across not only the acoustic conditions (“take#spins” vs. “takes#pins”) but also the recording conditions (concatenation vs. natural allophony), we measured target-detection latencies from the middle of the stable part of the target nucleus—for example, /I/ in “#spins” or “s#pins” in all conditions and for both target types (e.g., “spins” and “pins”). This location was defined operationally as the peak of the pitch period halfway into the stable cyclic portion of the vowel in case of an odd number of pitch periods or at the zero-crossing point of that period in case of an even number of pitch periods. All measures were made with CoolEdit Pro (Version 2.0) and the Praat speech editor (Boersma & Weenink, 2006). Using the middle of the nucleus as a reference point for measuring latencies introduced less arbitrariness than using the onset of the target would have, because the former was relatively unaffected by our acoustic and recording conditions. Moreover, the latter would have involved comparing the onsets of the long targets (always /s/) with those of the short targets (various consonants or a vowel), thus adding heterogeneity to the criteria for latency measurement. For the sake of comparability, however, we report analyses from target onset in Appendix C.

Because the task was explicitly presented to the participants as a broad phonological match, as opposed to one focusing on fine phonetic details, a correct response was tallied whenever a participant made a *target-present* response to a phonological match between a target and the signal, regardless of the acoustic or syntactic conditions. Thus, detecting “spins” or “pins” in either “take#spins” or “takes#pins” was counted as correct. Correct responses 2 standard deviations from the mean (computed separately for each participant) were discarded. Altogether, the discarded responses amounted to 19% of the test trials in the concatenated condition (10% incorrect) and 18% in the natural-allophony condition (11% incorrect). Average detection latencies and accuracy levels are reported in Table 2. Latencies, collapsed across the two recording types, are plotted in Figure 1.

The latency results showed a clear advantage of target–signal acoustic alignment in the neutral condition, with faster target detection in phrases containing congruent acoustic cues (e.g., “spins” in “take#spins,” “pins” in “takes#pins”) than in phrases containing incongruent acoustic cues (e.g., “pins” in “take#spins,” “spins” in “takes#pins”). However, this pattern was largely suppressed when the phrases were appended to a plural context. Here, a clear advantage for the syntactically congruent target emerged (e.g., “spins” in both “those women take#spins” and “those women takes#pins”), which suggests that the syntactic expectations generated by the plural subject had not only a strong impact on segmentation but also an attenuating effect on the acoustic cues. However, a different pattern was found in the singular condition in that the presence of a singular subject did not occasion shorter detection latencies for the syntactically congruent target (“pins”). Instead, the singular condition was similar to the neutral condition. None of these effects were significantly affected by whether the acoustic word-boundary cues were generated via concatenation or natural allophony.

These findings were supported by an analysis of variance with recording (concatenation, natural allophony), context (neutral, singular, plural), acoustics (“take#spins,” “takes#pins”), and target (“spins,” “pins”) as the main factors. For clarity, we report only comparisons having direct bearing on the main research questions (unless they reached .05 significance by both subjects and items). An effect of context significant by subjects, $F_1(2, 176) = 7.27$, $p = .001$; $F_2(2, 28) = 2.09$, $p = .14$, showed that targets tended to be responded to more rapidly in the singular and plural conditions $F_1(1, 88) = 10.24$, $p < .005$; $F_2(1, 14) = 5.93$, $p < .05$, than in the neutral condition, $F_1(1, 88) = 12.80$, $p = .001$; $F_2(1, 14) = 2.09$, $p = .17$. Latencies in the singular and plural conditions did not differ from each other, $F_1(1, 88) < 1$; $F_2(1, 14) < 1$. The shorter latencies in the two syntactic conditions probably reflect the general advantage conferred by a preceding context compared with neutral—and shorter—phrases. More critical was the Acoustics × Target interaction, $F_1(1, 88) = 19.49$, $p < .001$; $F_2(1, 14) = 9.22$, $p < .01$, modulated by context, $F_1(2, 176) = 3.40$, $p < .05$; $F_2(2, 28) = 3.96$, $p < .05$ (neither interaction was significantly affected by recording in subjects and items analyses simultaneously). The Acoustics × Target × Context interaction showed the following: In the neutral condition, an Acoustics × Target interaction, $F_1(1, 88) = 22.37$, $p < .001$; $F_2(1, 14) = 15.82$, $p = .001$, indicated that participants’ detection latencies were shorter if the acoustic cues aligned with the target, with “spins” detected faster in “take#spins” than in “takes#pins,” $F_1(1, 88) = 9.59$, $p < .005$; $F_2(1, 14) = 11.65$, $p < .005$, and “pins” detected faster in “takes#pins” than in “take#spins,” $F_1(1, 88) = 15.38$, $p < .001$; $F_2(1, 14) = 11.65$, $p < .005$. This interaction was only significant by subjects when the phrases were preceded by a singular context, $F_1(1, 88) = 6.49$, $p = .01$; $F_2(1, 14) < 1$. The Acoustics × Target interaction virtually disappeared when the phrases were presented in a plural context, $F_1(1, 88) = 2.77$, $p = .10$; $F_2(1, 14) < 1$. As before, none of these patterns were significantly affected by recording. Thus, consistent with the hypothesis that sublexical segmentation cues tend to be outweighed by higher level knowledge,

Table 2
Word-Detection Latencies (in Milliseconds) From the Nucleus of the Target and Accuracy (Percentages Correct, in Parentheses) in Experiment 1

Acoustics and target	Context: Neutral	Singular	Plural
Concatenated condition			
take#spins			
spins	465 (97)	465 (98)	416 (90)
pins	596 (81)	538 (78)	537 (67)
takes#pins			
spins	547 (87)	495 (76)	485 (83)
pins	507 (95)	491 (95)	533 (90)
Natural-allophony condition			
take#spins			
spins	559 (92)	498 (91)	517 (89)
pins	592 (92)	534 (84)	597 (77)
takes#pins			
spins	614 (81)	576 (73)	513 (79)
pins	543 (95)	541 (93)	575 (90)

Note. # = an acoustic juncture.

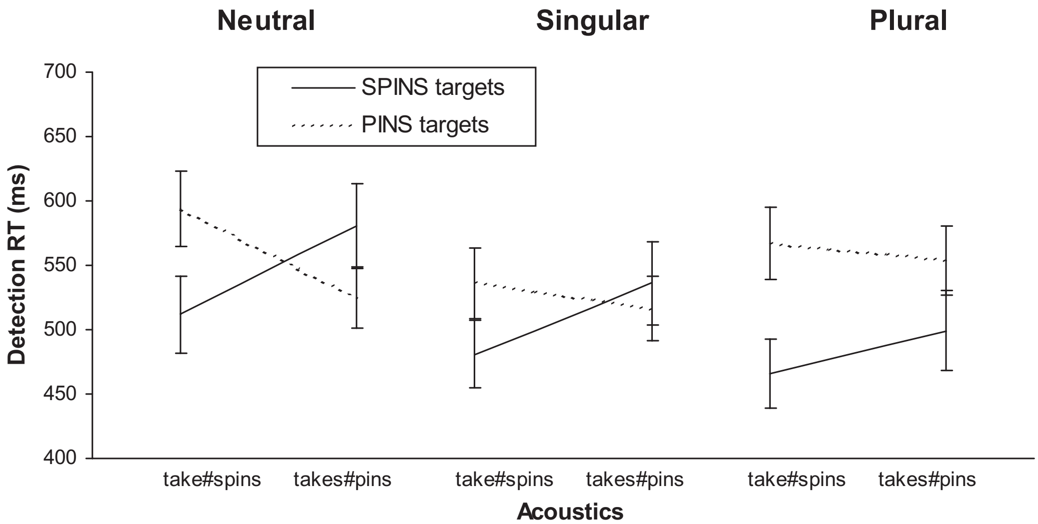

Figure 1. Target detection latencies (response times [RTs]), measured from the nucleus of the target, as a function of acoustic cues and syntactic context in Experiment 1 (with the concatenation and natural-allophony conditions collapsed). Error bars represent standard errors.

the syntactic expectations generated by the subject phrases caused the acoustic cues to be largely ignored.

We also investigated the more direct effect of syntax on segmentation—namely, the extent to which a singular subject sped up the detection of “pins” over “spins” and the extent to which a plural subject sped up the detection of “spins” over “pins.” A Context × Target interaction, $F_1(2, 176) = 11.24$, $p < .001$; $F_2(2, 28) = 3.05$, $p = .06$, suggested that the target effect did indeed differ across the three contexts. The neutral context did not show any significant bias for either target, $F_1(1, 88) = 1.13$, $p = .29$; $F_2(1, 14) = 1.92$, $p = .19$, whereas the plural context caused the expected advantage for “spins” over “pins,” $F_1(1, 88) = 37.23$, $p < .001$; $F_2(1, 14) = 10.06$, $p < .01$. However, the singular context did not result in an advantage for “pins” over “spins,” as would have been expected by the syntactic-segmentation hypothesis, $F_1(1, 88) = 1.92$, $p = .17$; $F_2(1, 14) = 4.79$, $p = .05$—if anything, the difference was in the opposite direction. None of these patterns were significantly affected by recording in subjects and items analyses simultaneously.

Analyses run on the accuracy data revealed similar trends, except that the Acoustics × Target interaction remained significant in all three context conditions: neutral, $F_1(1, 88) = 25.34$, $p < .001$; $F_2(1, 14) = 24.20$, $p < .001$; singular, $F_1(1, 88) = 51.79$, $p < .001$; $F_2(1, 14) = 46.00$, $p < .001$; and plural, $F_1(1, 88) = 25.73$, $p < .001$; $F_2(1, 14) = 56.78$, $p < .001$. Thus, participants missed the target more often when its onset was misaligned with an acoustic juncture. As in the latency results, a Context × Target interaction, $F_1(2, 176) = 6.09$, $p < .005$; $F_2(2, 18) = 9.17$, $p < .005$, suggested somewhat greater accuracy for “spins” (85%) than for “pins” (81%) in the plural condition, $F_1(1, 88) = 5.41$, $p < .05$; $F_2(1, 14) = 3.38$, $p = .09$, and no significant difference in the neutral, $F_1(1, 88) < 1$, $F_2(1, 14) = 2.11$, $p = .17$ (89% vs. 91%), or singular, $F_1(1, 88) = 2.30$, $p = .13$; $F_2(1, 14) = 1.96$, $p = .18$ (84% vs. 88%) conditions. None of these patterns were affected by recording (all $ps > .05$ by subjects and items).

Discussion

The results of Experiment 1 confirm that listeners use acoustic information to guide speech segmentation (e.g., Davis et al., 2002; Mattys, 2004; Mattys & Melhorn, in press; Salverda et al., 2003). It is important to note that there was no evidence that artificially inducing word-boundary junctures via concatenation enhanced the effect of acoustic cues on segmentation when compared with natural allophonic cues, although the specific acoustic cues leading to segmentation effects differed somewhat across the two conditions (see Appendix B for comparative analyses). Furthermore, the results indicate that the use of acoustic cues for segmentation can be modulated by syntactic expectations, especially in terms of response latency. However, contrasting results were found in the singular and plural conditions: Whereas a plural subject caused syntactically led segmentation, a singular subject did not.

A reason for this imbalance could be that the singular parsing (“takes pins”) was accidentally semantically less adequate for the subject phrase than was the plural parsing (“take spins”). That is, although the subject phrases (“that woman” and “those women”) were chosen to be semantically neutral relative to the two possible verb phrases (“take spins” and “takes pins”), the meaning of the subject phrases could have been unintentionally more strongly associated with the plural-consistent verb phrase than with the singular-consistent verb phrase. If so, the syntactic expectations generated by the singular subject would have had to compete not only against inconsistent acoustics but also against inconsistent semantics.

We addressed the issue of syntactic and semantic expectations in Experiment 2. First, we gathered completion ratings on the sentence fragments. Second, we formally aligned semantics with syntactic expectations. Thus, in this case, we chose the singular subject to be semantically related to the singular-consistent verb phrase (e.g., “The new machine at the bowling alley takes pins”) and the plural subject to be semantically related to the plural-consistent verb phrase (e.g., "Boys like it when their go-karts take spins"). In doing so, we gave syntactic expectations a better chance to show a modulatory effect on the processing of acoustic cues for segmentation.

Experiment 2

Method

Participants. In the main part of the experiment, 90 native speakers of British English, with the same characteristics as those in Experiment 1, were randomly assigned to either the concatenated ($n = 45$) or the natural-allophony ($n = 45$) condition.

Materials, design, and procedure. These were the same as in Experiment 1, except that new sets of singular and plural subject phrases were recorded. Each singular subject phrase was chosen to be semantically more strongly associated with the singular-consistent verb–object phrase than with the plural-consistent phrase, and vice versa. Selection of the subject phrases was done as carefully as possible within the constraints imposed by the existing test verb–object phrases. A set of stimuli is displayed in Table 1. All test materials are listed in Appendix D.

In an attempt to verify that the new subject phrases created the intended bias, we submitted them, along with the subject phrases of Experiment 1, to a completion-rating study. The stimuli consisted of a written version of the 30 subject phrases from Experiment 1 (*That woman* . . ., *Those women* . . . × 15) and the 30 new subject phrases from Experiment 2 (*The new machine at the bowling alley* . . ., *Boys like it when their go-karts* . . . × 15). One hundred and sixty participants, not used in any of the other experiments, were asked to rate the likelihood that the sentence would end with one alternative (e.g., *takes spins*) or the other (e.g., *takes pins*) on an 11-point scale. For each sentence, subject–verb number agreement was adjusted to be syntactically correct. For instance, the ending of *That woman* . . . was rated between *takes spins* and *takes pins*, whereas the ending of *Those women* . . . was rated between *take spins* and *take pins*. Sentences from Experiments 1 and 2 were mixed, but participants were split into four groups such that each participant never saw the same pair of alternatives twice. Thus, each participant rated 15 sentences, 7 from Experiment 1 and 8 from Experiment 2 (or vice versa). Furthermore, participants were assigned to one of two random orders, with the left–right position of the two alternatives counterbalanced as well. The 15 sentences were presented on a single sheet of paper, with a subject phrase and its two possible endings (separated by an 11-point scale) placed on two consecutive lines. Ratings were subsequently given values ranging from 0 for the phrase with the longer phrase-final word (*spins*) to 1 for the phrase with the shorter phrase-final word (*pins*).

The ending of the singular subject phrases of Experiment 1 obtained an average rating of .48, whereas those of the plural phrases received an average of .50, $t(14) = -.87$, $p = .40$. Neither condition departed significantly from the midpoint value .50, $t(14) = -.25$, $p = .80$, and $t(14) = .03$, $p = .97$, respectively. Thus, the subject phrases in Experiment 1, in fact, turned out to be relatively neutral with respect to the meaning of both potential endings. The subject phrases in Experiment 2, in contrast, showed the intended semantic bias. The ending of the singular subject phrases (e.g., *The new machine at the bowling alley* . . .) obtained an average rating of .17, showing a bias for the semantics of *takes pins*, whereas the ending of the plural subject phrases (e.g., *Boys like it when their go-karts* . . .) obtained an average rating of .85, showing a bias for the semantics of *take spins*. Both scores departed significantly from the midpoint value .50, $t(14) = -13.60$, $p < .001$, and $t(14) = 11.75$, $p < .001$, respectively. The distance between these scores and .50 was comparable, $t(14) = -.39$, $p = .70$. Finally, ratings of the singular subject phrases across experiments (.48 vs. .17) were indeed different, $t(14) = 5.55$, $p < .001$, as were those of the plural subject phrases (.50 vs. .85), $t(14) = -6.54$, $p < .001$. Thus, the stimuli of Experiment 2 significantly and symmetrically supplemented the syntactic cues of Experiment 1 with semantic information.

The new subject phrases for Experiment 2 were recorded by the speaker of Experiment 1 and appended to the beginning of the original test phrases, following the procedure described above. As before, participants were instructed to detect prespecified targets (e.g., "spins," "pins") in test phrases played in isolation (e.g., "take#spins," "takes#pins"), preceded by the singular subject phrase (e.g., "The new machine at the bowling alley take#spins"/ ". . . takes#pins"), and preceded by the plural subject phrase (e.g., "Boys like it when their go-karts take#spins"/". . . takes#pins").

Results

Incorrect responses and response latencies beyond the 2-standard deviation cutoff amounted to 20% of the test trials in the concatenated condition (15% incorrect) and 23% of the test trials in the natural-allophony condition (17% incorrect). Average detection latencies and accuracy are reported in Table 3. Latencies, collapsed across recording types, are plotted in Figure 2.

The latency results were similar to those in Experiment 1 in most respects. In the neutral condition, segmentation was influenced by the acoustic cues, whereas in the plural condition, it

Table 3
Word-Detection Latencies From the Nucleus of the Target (in Milliseconds) and Accuracy (Percentages Correct, in Parentheses) in Experiment 2

Acoustics and target	Context: Neutral	Singular	Plural
Concatenated condition			
take#spins			
spins	441 (95)	389 (81)	323 (90)
pins	531 (87)	469 (83)	459 (72)
takes#pins			
spins	498 (87)	454 (80)	349 (79)
pins	483 (97)	422 (87)	482 (84)
Natural-allophony condition			
take#spins			
spins	417 (95)	347 (81)	296 (89)
pins	480 (89)	414 (85)	456 (75)
takes#pins			
spins	463 (80)	395 (67)	376 (79)
pins	432 (94)	374 (81)	456 (87)

Note. # = an acoustic juncture.

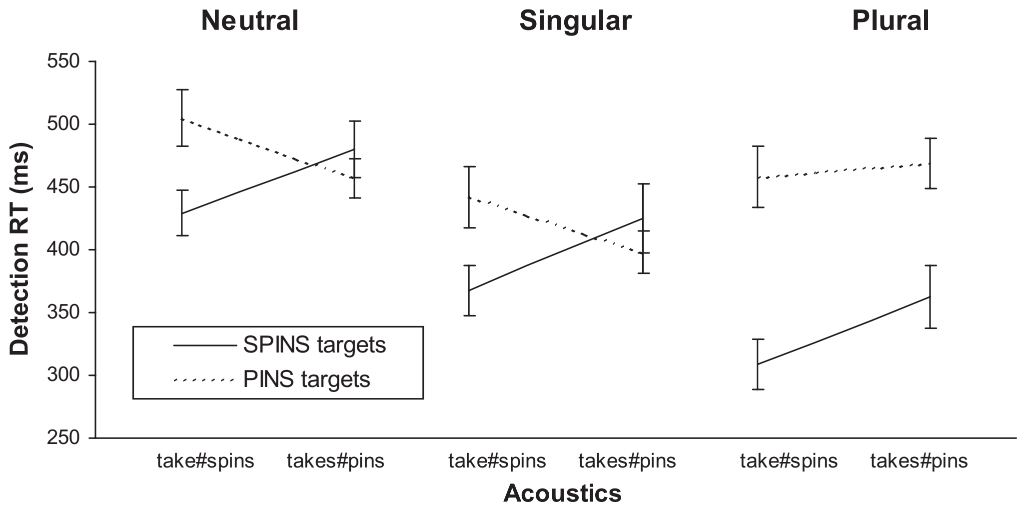

Figure 2. Target detection latencies (response times [RTs]), measured from the nucleus of the target, as a function of acoustic cues and syntactic context in Experiment 2 (with the concatenation and natural-allophony conditions collapsed). In this experiment, the syntactic context was supplemented with semantic cues. Error bars represent standard errors.

followed the syntactic (and semantic) context. The singular condition, as before, did not show any evidence of a syntactic effect, despite being reinforced by semantic information. Instead, it showed substantial sensitivity to the acoustic information, similar to the neutral condition. None of these effects were significantly affected by whether the acoustic cues were generated via concatenation or natural allophony.

An analysis of variance with the same factors as in Experiment 1 showed an effect of context, $F_1(2, 176) = 30.06, p < .001$; $F_2(2, 28) = 8.76, p = .001$, with targets responded to more rapidly in the singular and plural conditions, $F_1(1, 88) = 50.76, p < .001$; $F_2(1, 14) = 8.80, p = .001$, than in the neutral condition, $F_1(1, 88) = 38.44, p < .001$; $F_2(1, 14) = 14.60, p < .005$. Latencies in the singular and plural conditions did not differ from each other, $F_1(1, 88) < 1$; $F_2(1, 14) = 1.39, p = .26$. As in Experiment 1, we found an Acoustics × Target interaction, $F_1(1, 88) = 22.00, p < .001$; $F_2(1, 14) = 3.57, p = .08$. Although it was not significantly modulated by context, $F_1(2, 176) = 2.15, p = .12$; $F_2(2, 28) = 1.60, p = .22$, planned comparisons revealed contrasts similar to those found in Experiment 1: In the neutral condition, and to a lesser extent in the singular condition, an Acoustics × Target interaction indicated that detection latencies were influenced by the acoustic alignment between targets and the signal (neutral: $F_1[1, 88] = 18.44, p < .001$; $F_2[1, 14] = 7.85, p < .05$; singular: $F_1[1, 88] = 11.82, p = .001$; $F_2[1, 14] = 1.55, p = .23$). In particular, "spins" was detected faster in "take#spins" than in "takes#pins," at least by subjects (neutral: $F_1[1, 88] = 11.16, p = .001$; $F_2[1, 14] = 2.24, p = .16$; singular: $F_1[1, 88] = 7.49, p < .01$; $F_2[1, 14] = 5.09, p < .05$), and "pins" was detected faster in "takes#pins" than in "take#spins" (neutral: $F_1[1, 88] = 10.27, p < .005$; $F_2[1, 14] = 3.87, p = .07$; singular: $F_1[1, 88] = 6.04, p < .05$; $F_2[1, 14] < 1$). As before, the singular condition did not show any advantage for the syntactically consistent target ("pins" over "spins"), $F_1(1, 88) = 2.31, p = .13$; $F_2(1, 14) = 2.27, p = .15$. The plural condition, in contrast, caused the Acoustics × Target interaction to virtually disappear, $F_1(1, 88) = 3.27, p = .07$; $F_2(1, 14) < 1$, and a clear advantage for plural-consistent targets to emerge ("spins" over "pins"), $F_1(1, 88) = 73.68, p < .001$; $F_2(1, 14) = 35.97, p < .001$. The latter effect was significantly larger than that in the neutral condition, $F_1(1, 88) = 22.24, p < .001$; $F_2(1, 14) = 15.18, p < .005$, and that in the singular condition, $F_1(1, 88) = 37.39, p < .001$; $F_2(1, 14) = 12.86, p < .005$.

Finally, an analysis combining Experiments 1 and 2 indicated that the target effect in the plural condition was larger in Experiment 2 than Experiment 1, at least in the subjects analyses, $F_1(1, 176) = 9.61, p < .005$; $F_2(1, 14) = 2.52, p = .13$. Thus, there was some evidence that the added semantic information was efficient in further promoting reliance on syntax in the plural condition. The added information did not have any noticeable effect in the singular condition (all *p*s > .05). All other comparisons across the two experiments failed to reach significance at the $p < .05$ level by both subjects and items. However, a main effect of experiment, $F_1(1, 176) = 11.66, p = .001$; $F_2(1, 14) = 205.57, p < .001$, indicates that respondents were faster in Experiment 2 than in Experiment 1. Although part of this effect is probably attributable to the longer singular and plural sentences in Experiment 2, the generalization of this effect to the neutral conditions, which were identical in both experiments, suggests sampling disparities between the two groups—with no obvious cause, but possibly attributable to the time in the academic year when participants were tested.

The accuracy analyses showed some (albeit limited) sensitivity to the added semantic information for the singular condition. The Acoustics × Target interaction was modulated by context, $F_1(2, 176) = 5.94, p < .005$; $F_2(2, 28) = 3.12, p = .06$, in that it was significant in the neutral condition, $F_1(1, 88) = 22.59, p < .001$; $F_2(1, 14) = 33.42, p < .001$, and in the plural condition, $F_1(1, 88) = 30.44, p < .001$; $F_2(1, 14) = 24.69, p < .001$, but disappeared in the singular condition, $F_1(1, 88) = 2.79, p = .10$; $F_2(1, 14) = 1.73, p = .21$. A Context × Target interaction, $F_1(2,$

176) = 11.77, $p < .001$; $F_2(2, 18) = 8.96$, $p < .005$, indicated slightly greater accuracy for "pins" than for "spins" in the neutral condition (92% vs. 89%), $F_1(1, 88) = 2.78$, $p < .10$; $F_2(1, 14) = 6.03$, $p < .05$, and in the singular condition (84% vs. 77%), $F_1(1, 88) = 8.99$, $p < .005$; $F_2(1, 14) = 8.10$, $p < .05$, and greater accuracy for "spins" than for "pins" in the plural condition (84% vs. 79%), $F_1(1, 88) = 7.60$, $p < .001$; $F_2(1, 14) = 8.73$, $p = .01$. However, although the target effect in the plural condition clearly differed from that in the neutral condition, $F_1(1, 88) = 12.19$, $p < .001$; $F_2(1, 14) = 14.11$, $p < .005$, that in the singular condition did not do so to a fully reliable extent, $F_1(1, 88) = 4.07$, $p < .05$; $F_2(1, 14) < 1$. None of the above latency or accuracy patterns were affected by recording (all ps > .05 by subjects and items).

Discussion

On the whole, the results indicate that supplementing syntactic expectations with consistent semantic information did not allow a clear syntactic effect to emerge in the singular condition. As in Experiment 1, the plural subject phrases were efficient in causing syntax-sensitive segmentation, and some benefit from the added semantics was observed in the latency results, although this was only moderate. Thus, the imbalance between the singular and plural conditions noted in Experiment 1 was probably not caused by a bias in semantic information. Further possibilities are explored in the next section.

Acoustic and Time-Course Considerations

Another explanation for the asymmetry between the singular and plural results could be sought at the acoustic level. Because of the nature of the experimental design, the onset segment of the plural-consistent targets was always /s/, whereas the onset of the singular-consistent targets varied from target to target (/p/, /t/, /n/, /m/, /l/, or /i/). Given the phoneme-specific nature of allophonic cues for word juncture (Nakatani & Dukes, 1977; Nakatani & Schaffer, 1978), it is possible that the singular-inconsistent acoustic cues ("take#spins") could have been stronger than the plural-inconsistent cues ("takes#pins"), thus making an effect of syntactic expectations less likely to manifest in the singular than in the plural condition. If so, this imbalance should be reflected in the detection results of the neutral conditions as well. If the "take#spins" phrases did indeed contain stronger acoustic cues than the "takes#pins" phrases did, then the target effect in the "take#spins" phrases in the neutral context should have been larger than that in the "takes#pins" phrases. We calculated an *acoustic advantage* index for the neutral conditions of Experiments 1 and 2 as the difference between the detection latency for the acoustically congruent target and that for the acoustically incongruent target for each condition of the following design: acoustics ("take#spins," "takes#pins") and recording (concatenation, natural allophony). (Whether the data originated from Experiments 1 or 2 was not entered in the analysis because the neutral condition was identical for both experiments.) The same was done for accuracy. The latency results indicated that the singular-inconsistent phrases ("take#spins") were indeed more conducive to segmenting the appropriate target than were the plural-inconsistent phrases ("takes#pins"), $F_1(1, 176) = 6.55$, $p = .01$; $F_2(1, 14) = 4.88$, $p < .05$, with acoustic advantages of 79 ms and 39 ms, respectively. However, a slightly opposite trend was observed in the accuracy data, $F_1(1, 176) = 3.05$, $p = .08$; $F_2(1, 14) = 5.04$, $p = .04$ (7% vs. 11%, respectively).

Thus, there was some evidence that the asymmetric effect of syntax in the singular and plural conditions could have been a result of an imbalance in the strength of the critical acoustic cues. To test this hypothesis further, we reran the analyses in Experiments 1 and 2 with a subset of more acoustically balanced stimuli. To do so, we rank ordered the 15 sets of stimuli according to the size of their acoustic-advantage asymmetry (in terms of latency), averaged across the neutral conditions of Experiments 1 and 2. Of the 15 sets, we kept only the 10 sets with the smallest acoustic-advantage asymmetry. The acoustic advantages in the "take#spins" and "takes#pins" phrases, after stimulus selection, were 61 ms and 63 ms, respectively, $F_1(1, 176) < 1$; $F_2(1, 9) < 1$. An analysis of variance was then run on the new set of stimuli, with context (neutral, singular, plural), acoustics ("take#spins," "takes#pins"), and target ("spins," "pins") as the main factors. Recording (concatenation, natural allophony) and experiment (1, 2), although entered in the analysis, are not discussed here because they were not found to substantially modulate the asymmetry between the singular and plural contexts in the earlier analyses. The overall pattern of results (see Figure 3) was not notably different from that in the first two experiments. A Context × Acoustics × Target interaction, $F_1(2, 352) = 4.32$, $p = .01$; $F_2(2, 18) = 4.37$, $p < .05$, showed an impact of the acoustic cues (Acoustics × Target interaction) in the neutral and singular conditions (neutral: $F_1[2, 176] = 37.66$, $p < .001$; $F_2[1, 9] = 9.82$, $p = .01$; singular: $F_1[2, 176] = 14.09$, $p < .001$; $F_2[1, 9] = 6.05$, $p < .05$), but less consistently so in the plural condition, $F_1(2, 176) = 5.65$, $p = .02$; $F_2(1, 9) = 1.47$, $p = .26$.

The target effect also differed across context conditions, $F_1(2, 352) = 22.07$, $p < 001$; $F_2(2, 18) = 5.58$, $p = .01$. The neutral context did not show any significant bias for either target, $F_1(1, 176) < 1$; $F_2(1, 9) < 1$. As before, the plural context caused a syntactically consistent advantage for "spins" over "pins" targets, $F_1(1, 176) = 63.70$, $p < .001$; $F_2(1, 9) = 12.44$, $p < .01$, but the mirror effect was not found in the singular condition. If anything, "spins" targets were slightly easier to detect than "pins" targets, $F_1(1, 176) = 4.51$, $p < .05$; $F_2(1, 9) = 3.71$, $p = .09$. Thus, these results disconfirm the hypothesis that the absence of a syntactic effect in the singular condition of Experiments 1 and 2 was a result of acoustic cues that were harder to override in the "take#spins" condition than in the "takes#pins" condition.

The imbalance between the singular and plural conditions could also be attributable to a difference in the relative time of arrival between the acoustic and syntactic information. In the singular context, the discrepant acoustic cue occurred before the verb inflexion is realized ("That woman take#s]pins" [with the bracket indicating the syntactically consistent segmentation point]), whereas in the plural case, it occurred after it ("Those women take]s#pins"). Thus, from a sequential viewpoint, commitment to syntax in the singular condition would require that listeners ignore the intervening conflicting acoustic cues or at least suppress the activation of the lexical candidates with which the cues align. Commitment to syntax in the plural condition would not suffer from this constraint as much.

If syntactic effects depend on whether conflicting acoustic cues are encountered prior to syntactic realization, then the size of the

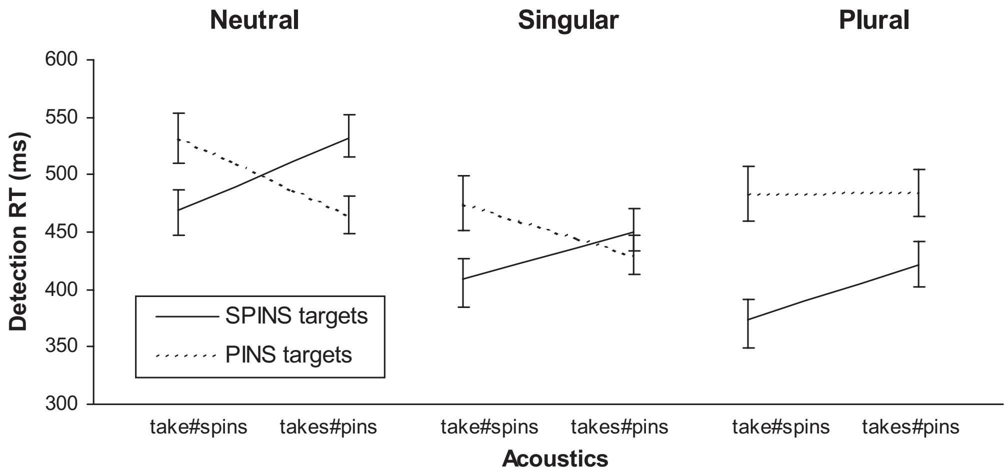

Figure 3. Collapsed results of Experiments 1 and 2 with only the 10 most acoustically symmetrical stimulus sets (out of 15). Error bars represent standard errors. RT = response time.

syntactic effect and that of the acoustic effect should trade off differently in the singular and plural conditions. In the singular condition, in which discrepant acoustics are encountered before syntactic realization, the size of the syntactic effect should be inversely related to that of the acoustic effect—the milder the acoustic cues, the larger the syntactic effect. In the plural condition, however, the size of the syntactic effect should be relatively independent of that of the acoustic cues, because commitment to syntax can take place before the conflicting acoustic cues are encountered. Correlational analyses revealed the expected contrast. The acoustic effect size of each of the 15 stimulus sets was estimated as the difference between the average latency of the two acoustically incongruent conditions ("pins" in "take#spins" and "spins" in "takes#pins") and the average latency of the two acoustically congruent conditions ("spins" in "take#spins" and "pins" in "takes#pins"). The syntactic effect size was measured as the difference between the latencies to the syntactically incongruent target—regardless of the acoustics—and the latencies to the syntactically congruent target. This was done separately for the singular and plural conditions. Recording (concatenation, natural allophony) and experiment (1, 2) were, once again, collapsed. Scatterplots of the 15 stimulus sets are shown in Figure 4. Although a comparison between the two correlation coefficients using the *ZPF* statistic (Raghunathan, Rosenthal, & Rubin, 1996) did not reach significance, there was evidence of a contrast between the singular and plural conditions when these were considered separately. In the singular condition, we found a trade-off pattern between acoustic and syntactic effects in which the smaller the acoustic effect, the larger the syntactic effect ($r = -.55$, $p = .03$). The plural condition did not show any significant correlation ($r = -.06$, $p = .82$). In an attempt to factor out any intrinsic bias in the phrases themselves, we reran these analyses controlling for

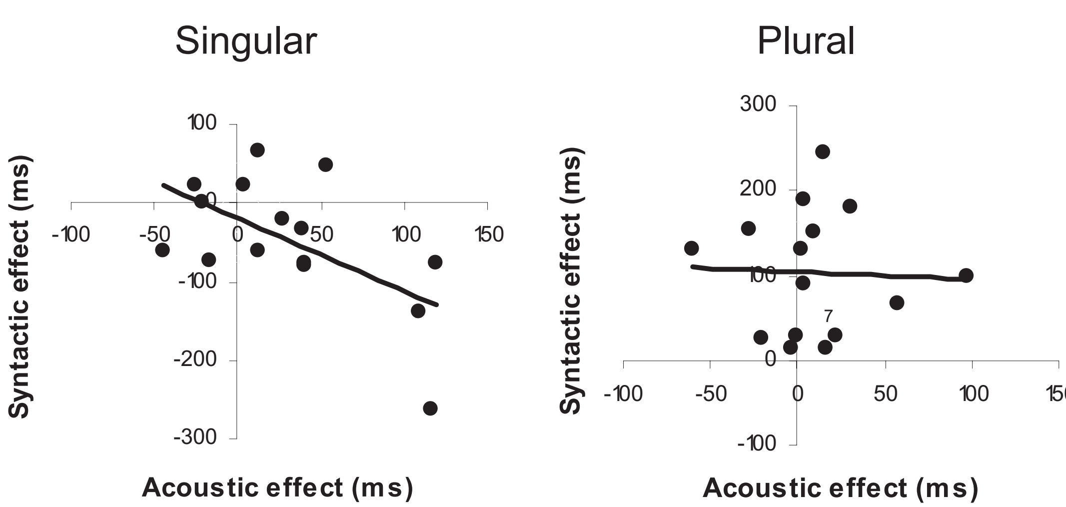

Figure 4. Scatterplots of the acoustic versus syntactic effect sizes for the 15 stimulus sets in the singular and plural conditions of Experiments 1 and 2 (collapsed). The singular condition shows a trade-off relation between acoustic and syntactic effects, whereas the plural condition shows none.

the acoustic effect size and the relevant syntactic effect size in the neutral condition. The results were unchanged (singular, $r = -.73$, $p = .005$; plural, $r = .22$, $p = .47$).

Although the above analyses were post hoc, they suggest that the relative time of arrival of acoustic cues and syntactic realization affects reliance on the two sources of information, with the earlier occurring cue being given priority. Thus, if the effect of syntactic expectations in the singular condition was indeed hindered by the intervening acoustic cues, then decreasing the strength of the acoustic cues should allow the syntactic effect to emerge—as suggested by the negative correlation. Therefore, we tested this hypothesis with a set of neutral recordings (piloted in Experiment 3A).

Experiment 3A: Stimulus Selection

The goal of Experiment 3A was to generate a set of acoustically neutral renditions of the phrases used in Experiments 1 and 2 (e.g., "takespins" yielding comparable "take spins" and "takes pins" perceptual judgments). A large number of acoustically neutral recordings were made and then rated by participants. The most neutral rendition for each of the 15 original phrases was kept for Experiment 3B.

Method: Materials, Recordings, Participants, and Procedure

The speaker of Experiments 1 and 2 produced five renditions for each of the 15 test phrases. He was asked to pronounce the phrases such that they were as acoustically ambiguous as possible (e.g., between "take spins" and "takes pins"). The 75 phrases (5×15) were then played to 30 native speakers of British English. The phrases were presented in a different random order for each participant. Participants were asked to rate each rendition using an 11-point scale ranging from one possible interpretation ("take spins") to the other ("takes pins"). On each trial, a phrase was played and then immediately followed by the two written alternatives on a computer monitor. The two alternatives were separated by a series of dots. Participants used eleven designated keys on a computer keyboard to indicate their response. The left–right position of the two alternatives was counterbalanced between phrases and participants. On the push of a response key, a 1-s interval elapsed before the next phrase was played.

Results

Ratings were recoded from .00 (e.g., "take spins") to 1.00 (e.g., "takes pins") in steps of .10, with .50 corresponding to an answer equidistant from both extremes. The most centrally rated rendition of each test phrase was kept, yielding an average of .52.

Experiment 3B

This experiment was similar to Experiments 1 and 2, except that the test phrases were replaced with the acoustically neutral renditions selected in Experiment 3A.

Method

Participants. Ninety native speakers of British English were randomly assigned to either a syntax-only ($n = 45$; cf. Experiment 1) or a syntax–semantics ($n = 45$; cf. Experiment 2) condition.

Materials, design, and procedure. All materials but the test phrases were the same as in Experiments 1 and 2. The acoustically neutral test phrases selected in Experiment 3A were concatenated to the subject phrases of Experiment 1 (which provided a syntactically biasing context) and to the subject phrases of Experiment 2 (which provided a syntactically and semantically biasing context). Participants heard either the syntax-only utterances or the syntax–semantics utterances. Within each group, context conditions (neutral, singular, plural) were rotated across stimulus sets and participants following the Latin square design described in the previous experiments. Because the test phrases had only one acoustic realization, the number of test trials in this experiment was half that in the previous experiments. The filler utterances were the same as in Experiments 1 and 2. The digital concatenation of the subject phrases and the test phrases followed the same procedure as in Experiment 1. As before, on each trial, a visual target (e.g., *spins*) appeared on a computer monitor prior to the beginning of the utterance.

Results

Incorrect responses and response latencies beyond the 2-standard deviation cutoff amounted to 9% of the test trials in the syntax-only condition (7% incorrect) and 9% of the test trials in the syntax–semantics condition (5% incorrect). The higher accuracy in this experiment compared with Experiments 1 and 2 probably reflects the acoustic neutrality of the current test phrases, in which acoustic cues were not so strong as to inhibit the detection of either target type. Average detection latencies, measured from the onset of the target nucleus, and accuracy levels are reported in Table 4. Latencies are plotted in Figure 5, separately for the syntax-only and syntax–semantics conditions.

The results showed an effect of syntactic expectations in both the singular and plural conditions, independent of whether syntactic expectations were supplemented with semantic information. An analysis of variance performed on the detection latencies with group (syntax only, syntax–semantics), context (neutral, singular, plural), and target (*spins*, *pins*) as factors showed an effect of context, $F_1(2, 176) = 28.30$, $p < .001$; $F_2(2, 28) = 17.64$, $p < .001$. As in Experiments 1 and 2, targets were responded to more rapidly in the singular and plural conditions, $F_1(1, 88) = 10.17$,

Table 4
Word-Detection Latencies From the Nucleus of the Target (in Milliseconds) and Accuracy (Percentages Correct, in Parentheses) in Experiment 3B

Target	Context: Neutral	Singular	Plural
Syntax only			
spins	583 (93)	527 (90)	432 (95)
pins	594 (97)	472 (93)	532 (91)
Syntax and semantics			
spins	654 (94)	567 (92)	444 (96)
pins	644 (97)	499 (97)	585 (92)

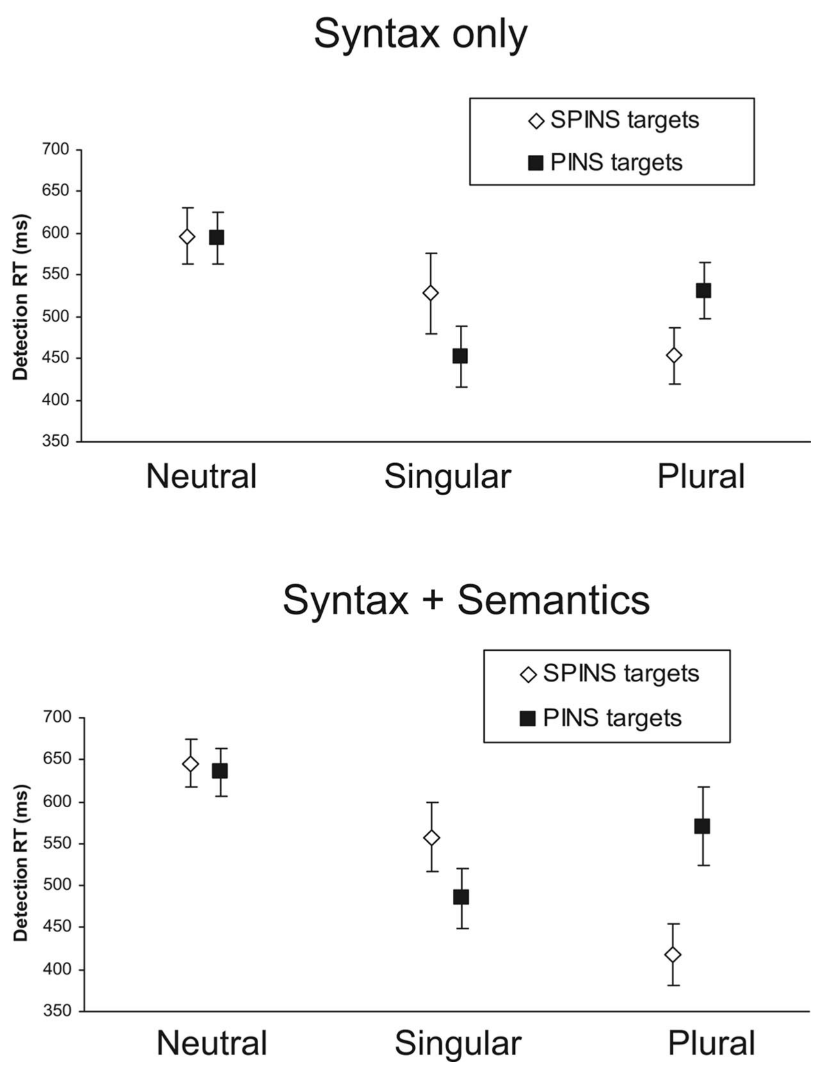

Figure 5. Target detection latencies (response times [RTs]), measured from the nucleus of the target, as a function of the acoustic context in Experiment 3B. The top panel displays the results when the contextual information is restricted to syntactic expectations (cf. Experiment 1). The bottom panel displays the results when the contextual information includes both syntactic and semantic information (cf. Experiment 2). Error bars represent standard errors.

$p < .01$; $F_2(1, 14) = 20.50$, $p < .001$, than in the neutral conditions, $F_1(1, 88) = 17.41$, $p < .001$; $F_2(1, 14) = 35.86$, $p < .001$. Latencies in the singular and plural conditions did not differ from each other, $F_1(1, 88) = 2.03$, $p = .16$; $F_2(1, 14) < 1$. More critically, a Context × Target interaction, $F_1(2, 176) = 24.68$, $p < .001$; $F_2(2, 28) = 19.12$, $p < .001$, indicated equal latencies for the "spins" and "pins" targets in the neutral condition, $F_1(1, 88) < 1$; $F_2(1, 14) < 1$, shorter latencies for "pins" than for "spins" targets in the singular condition, $F_1(1, 88) = 12.98$; $p = .001$; $F_2(1, 14) = 4.75$, $p < .05$, and the opposite in the plural condition, $F_1(1, 88) = 37.30$, $p < .001$; $F_2(1, 14) = 16.67$, $p = .001$. Thus, syntactically congruent targets were detected faster than syntactically incongruent targets in both the singular and plural conditions. None of these effects were significantly affected by the group factor. Although numerically comparable, none of the analyses carried out on the accuracy results revealed effects or interactions significant by both subjects and items.

To test whether the singular and plural conditions differed in the magnitude of their effect on segmentation, we compared the size of the syntactic effect for the singular and plural conditions. The overall syntactic effect, which can be inferred from Table 4, was 55 ms (singular) and 68 ms (plural) for the syntax-only group, and 100 ms (singular) and 141 ms (plural) for the syntax–semantics group. An analysis of variance run on the size of the syntactic effect, with group (syntax-only, syntax–semantics) and context (singular, plural) as independent variables, did not show a group effect, $F_1(1, 88) = 1.78$, $p = .18$; $F_2(1, 14) = 1.28$, $p = .28$, a context effect, $F_1(1, 88) = 2.03$, $p = .16$; $F_2(1, 14)$ 1.46, $p = .25$, or a Group × Context interaction, $F_1(1, 88) = 1.90$, $p = .17$; $F_2(1, 14) < 1$. Therefore, the magnitude of the syntactic effect was comparable in the singular and plural conditions, and there was no statistical evidence that the added semantics benefited segmentation over and above the advantage provided by syntax.

Discussion

This experiment showed a clear effect of syntactic expectations on word-monitoring latencies: Target words that aligned well with a syntactically consistent boundary were detected faster than were those that did not. In contrast to the earlier experiments, in which syntactic effects were found only in the plural condition, the present experiment showed effects for both the singular and plural contexts. The absence of a syntactic effect in the singular condition of Experiments 1 and 2 was attributed to the disrupting effect of conflicting acoustic cues prior to syntactic realization. Removing such cues in Experiment 3B did indeed cause a syntactic effect to emerge in the singular condition. Thus, as already suggested by the correlational analyses graphed in Figure 4, the relative time course of syntactic and acoustic cues seems to strongly affect the extent to which these cues are used for speech segmentation.

It should be noted that our attempt to understand how listeners dealt with a conflict between syntactic expectations and acoustic cues overlooked an important factor—namely, the *segmental* consequence of a misalignment between the two sources of information. By design, all of our experiments (including Experiment 3B) contained an interesting asymmetry in that respect. In the singular condition, a misalignment between syntax and the target implied that the pivotal /s/ had to be interpreted as both the verb inflection and the target's onset ("That woman take*spins*" [with the syntactically consistent verb underlined and the target in italics]). In contrast, a misalignment in the plural condition implied that the /s/ was lexically unassigned ("Those women takes*pins*). Research suggests that the former type of conflict might be less taxing than the latter. Indeed, fricative gemination, which would in principle be required to satisfy the conflict encountered in the singular condition ("takes spins"), is generally only realized as lengthened frication rather than full reduplication of the segment (e.g., for Italian, Giovanardi & Di Benedetto, 1998; see also Ladefoged, 2001). Thus, the syntax–target misalignment in the singular condition can be solved at a relatively low perceptual cost if the duration of the pivotal /s/ falls within an acceptable range for gemination. In contrast, the stranded /s/ in the plural condition is less likely to go unnoticed, not only because it involves disregarding a rather long and acoustically salient portion on the signal, but also because the isolated consonant violates possible-word constraints shown to be critical for segmentation (Norris, McQueen, Cutler, & Butterfield, 1997).

Although the analysis of variance in Experiment 3B did not show significant differences between the singular and plural conditions, correlational analyses between the duration of the pivotal /s/ and the size of the syntactic effect confirmed the impact of segmental constraints on syntax-sensitive segmentation. Duration of /s/ was measured for each of the 15 test phrases as the interval between the onset and the offset of the stable portion of the consonant's friction. As can be seen in Figure 6, long /s/ segments were associated with a small syntactic effect in the singular condition ($r = -.51$, $p < .005$; syntax only: $r = -.52$, $p < .05$; syntax–semantics: $r = -.50$, $p = .06$) but with a large syntactic effect in the plural condition ($r = .52$, $p < .005$; syntax only: $r = .45$, $p = .09$; syntax–semantics: $r = .60$, $p = .02$). The two correlations were found to be significantly different from each other, $t(27) = -4.29$, $p < .001$ (Meng, Rosenthal, & Rubin, 1992). Thus, in the singular condition, a misalignment between syntactic expectations and targets was more manageable if the critical /s/ was long, which suggests that perceived gemination can act as a compromise between syntax-sensitive and signal-sensitive segmentation. In contrast, the cost of a misalignment in the plural condition was greater with long /s/ sounds, as would be expected with a segmentation process that penalizes lexically unassigned segments. Thus, the effect of syntactic expectations on speech segmentation must be analyzed not only in the context of the strength of the competing acoustic cues or the relative time course of such cues but also in the context of the segmental cost that might result from syntax-sensitive segmentation.

General Discussion

How listeners segment words from fluent speech has so far been analyzed mainly at the lexical and sublexical levels. Few studies have examined the contribution of higher order information such as syntax, semantics, and pragmatics. In the present experiments, we set out to test whether syntactic expectations help listeners locate word boundaries and how syntactic expectations fare against conflicting sublexical cues. Although the experiments showed clear effects of syntactic expectations on speech segmentation, natural asymmetries in the design gave us an opportunity to examine important modulating factors.

Experiment 1 indicated that the parsing expectations promoted by a plural subject phrase outweighed conflicting acoustic cues, a result consistent with the hierarchy of cues proposed by Mattys White, and Melhorn (2005). However, a singular subject phrase, which was initially expected to mirror the plural condition, had no effect on segmentation. If anything, the results indicated dominance of the acoustic cues. Enhancing the syntactic expectations by adding semantic cues to the subject phrases did not significantly attenuate the singular–plural asymmetry (Experiment 2); it only slightly increased the syntactic effect in the plural condition. Likewise, although we subsequently found that the acoustic cues were not equally salient in the two conditions, controlling for this imbalance had no notable consequences on the initial pattern. It should be noted that none of these findings were significantly affected by whether the acoustic cues resulted from natural allophony or concatenation. However, this should not be seen as evidence that concatenated stimuli are the acoustic and perceptual equivalent of connected-speech allophony. The acoustic analyses shown in Appendix B did reveal qualitative or quantitative differences in how junctures were realized in the two conditions. Although the impact of these differences on perception was never manifest in higher level interactions, processing differences occasionally emerged in local comparisons. However, no systematic connections between mode of recording and complete segmentation patterns were found. It is therefore more parsimonious at this stage to interpret this result as showing generalization between the two modes of recording.

What emerged as a critical factor, however, was the relative time course of the acoustic cues and the realization of the syntactic expectations. In the singular condition, which showed acoustic dominance, the segmentation point suggested by the acoustic cues preceded that suggested by syntax ("That woman take#s]pins"), whereas in the

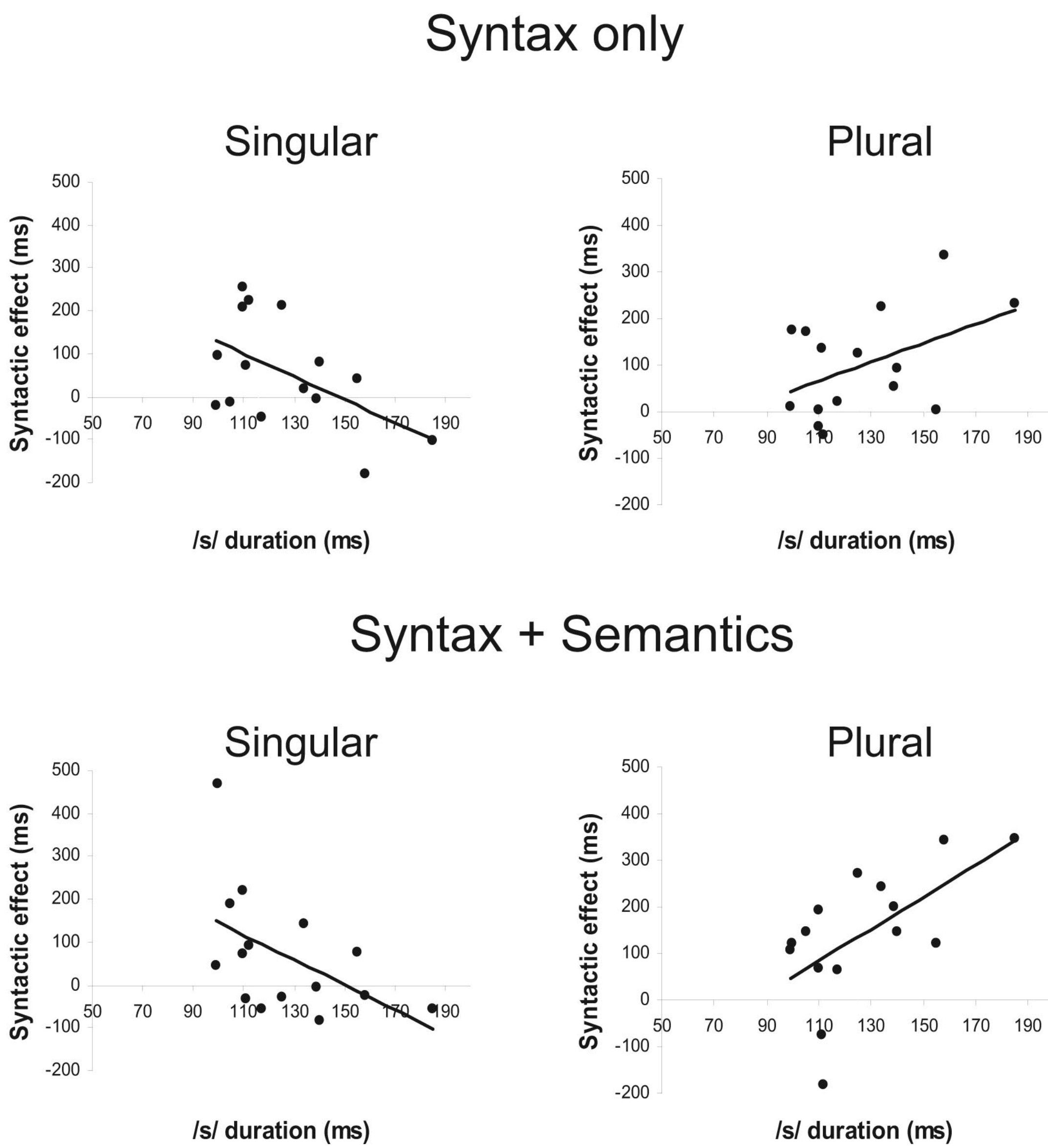

Figure 6. Scatterplots of the /s/ duration and the syntactic effect size for the 15 test phrases in the singular and plural conditions of Experiment 3B for the syntax-only and syntax + semantics conditions separately. The singular condition shows a negative correlation between /s/ duration and the syntactic effect, whereas the plural condition shows a positive correlation.

plural condition, it followed it ("Those women take]s#pins"). Correlational analyses highlighted the fact that the magnitude of the syntactic effect in the singular condition was inversely proportional to the size of the acoustic effect. No such correlation was found in the plural condition. This suggests that syntactic expectations are more effective when they are unimpeded by prior acoustic cues. Experiment 3 confirmed this hypothesis by highlighting that controlled neutralization of the acoustics allowed the syntactic effect to emerge in both the singular and plural conditions.

Finally, there were indications that reliance on syntactic knowledge was also contingent on its consequences for segmental plausibility. As suggested in the correlational analyses of Experiment 3, reliance on syntactic knowledge was less likely to be found if a conflict with the signal could be solved at a minimal perceptual cost (e.g., illusory gemination) than if it resulted in a severe segmental incongruity (e.g., lexically unassigned segments).

Taken together, the results suggest that syntactic knowledge has a sizeable impact on segmentation but that its interaction with sublexical cues is constrained by important variables. Mainly, syntax can exert its effect on segmentation as long as its realization (e.g., the actual verb inflection) occurs before any conflicting cues are encountered. Thus, the anticipatory effect of syntactic knowledge on segmentation is fragile and easily overridden by intervening conflicting cues. This apparently *late-commitment* feature of syntax-sensitive segmentation is not absolute, however, in that it appears to trade off with the strength of the intervening cues.

Similarly, compromises between syntax and the signal are also sought when their conflict yields a relatively mild segmental discrepancy. Thus, the relationship between syntactic knowledge and acoustic cues is more graded and dynamic than would be assumed by a strictly hierarchical model (for additional evidence on the notion of graded segmentation cues, see Mattys & Melhorn, in press). This is not entirely surprising given the radically different time course of these two types of information. Whereas acoustic cues operate locally (e.g., within a syllable's range), syntactic expectations generally have extended domains, building up over longer stretches of signal. Although models that have considered spoken-word recognition within a gradually unfolding signal have sometimes attempted to specify how acoustic cues constrain lexical activation (e.g., Davis et al., 2002; Gow & Gordon, 1995; Norris et al., 1997; Tabossi, Burani, & Scott, 1995), none have examined the interaction between syntactic constraints and lexical activation–selection. As an exception, Tyler and Wessels (1983) found that syntactic expectations somewhat reduced the amount of sensory information necessary to identify a gated spoken word, possibly via the deactivation of those word forms (e.g., nouns, verbs, adjectives) that are inconsistent with syntactic expectations.

Expanding on Tyler and Wessels's (1983) conclusion, our results suggest that syntactic expectations also exert their effect at the level of words' inflected forms, with candidates that align with syntactically consistent inflections being favored in the activation–segmentation process. This finding could be accommodated by a model that allows multiple inflected forms to be activated simultaneously, with levels of activation commensurate with the amount of prior syntactic evidence. The activation difference between syntactically consistent and inconsistent candidates would itself be influenced by whether these candidates align with acoustic cues. Contrary to a strict hierarchical approach, however, conflicts between syntax and acoustics would not necessarily be solved at the syntactic level (i.e., with syntax overriding acoustics). Which candidates retain the highest activation level would depend on (a) the time of arrival of the acoustic cues relative to the arrival of syntactic realization, (b) the strength of the acoustic cues in the sensory input, and (c) whether segmentation solutions leave lexically impossible residues.

Although the present data do not speak to whether inflected verb forms are stored as individual representations and enter lexical competition similarly to other words or whether they undergo an independent inflection process, the data suggest that distinct inflected forms—however they come about—can receive activation simultaneously in at least some circumstances. Multiple lexical activation models (e.g., Norris, 1994) could be amended to include syntactic constraints on lexical activation in addition to sublexical cues such as stress (Norris, McQueen, & Cutler, 1995) and phonetic–phonological cues (Norris et al., 1997). Here, the inflected forms of lexical candidates would receive activation levels in line with the unfolding syntactic evidence. However, the time course of such lexical activation–selection might differ from that elicited by sublexical cues. Whereas stress and phonetic–phonological cues are essentially properties of the activated candidates themselves, syntactic constraints are usually imposed by distal, earlier parts of the utterance. Thus, modulating the activity level of words that have not yet received any sensory input might require the involvement of short-term memory resources. This limitation could explain why the use of syntax for segmentation is fragile and contingent on time-course issues as well as intervening cues. In future experiments, one might want to assess the extent to which syntactic effects such as those that we observed are affected by a concurrent memory task and, conversely, whether performance on the memory task decreases as the distance between the subject phrase and the verb–object phrase increases. A moment-to-moment analysis of those candidates receiving most activation could also be carried out, for example, by using a cross-modal design in which targets are displayed at different points in time during sentence playback.

It should also be mentioned that in our effort to avoid some of the interpretive caveats attendant upon earlier studies, we restricted the scope of syntax to a narrow aspect—namely, subject–verb number agreement. In doing so, although we succeeded in isolating the influence of syntactic expectations on segmentation from their usual semantic correlates, we ignored other important aspects of syntactic knowledge such as noninflective grammatical relations (e.g., prepositional restrictions), word order, syntactic categories, tree structure, and so forth. It is likely that syntactic expectations, broadly defined, vary widely in terms of their constraining power on segmentation. However, it is also likely that the syntactic paradigm we used in this study lay toward the less constraining end of this spectrum. Indeed, our syntactic constraints involved single segments rather than larger, more salient units such as words or phrases. As syntactic constraints, segments are relatively flexible, not only because they vary in the extent to which they are realized (an acceptable verb inflection could be realized as little more than a brief frication) but also because the mandatory nature of verb inflection is, itself, subject to dialectal variations. For example, in some dialects of English, verbal inflections are consistent throughout the present tense paradigm (e.g., "I/you/he/they . . . like pins" or "I/you/he/they . . . likes pins"; see, e.g., Hughes & Trudgill, 1996). Yet, despite these only moderately constraining conditions, clear syntactic effects were observed. Thus, it is reasonable to assume that stricter syntactic constraints are also likely to have an effect on segmentation.

Finally, it is worth pointing out that the experiments in which semantic information was added to syntax only showed moderate benefits for segmentation. However, we do not take this result as an indication that the contribution of semantics to speech segmentation is minimal. Indeed, the semantic cues were created within the constraints of the existing materials simply to enhance the syntax contrast. As a result, many utterances provided only a partial semantic bias for a particular segmentation solution. A possibility for future research would be to pit syntax against semantics in a fully orthogonal design so as to tease out the unique contribution of semantics as well as to assess its interaction with syntax in cases of conflict.

References

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

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* (Release 2) [CD-ROM]. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Baum, S. R. (1989). On-line sensitivity to local and long-distance syntactic dependencies in Broca's aphasia. *Brain and Language, 37,* 327–338.

Baum, S. R. (1991). Sensitivity to syntactic violations across the age-span: Evidence from a word-monitoring task. *Clinical Linguistics and Phonetics, 5,* 317–328.

Boersma, P., & Weenink, D. (2006). Praat: Doing phonetics by computer (Version 4.4.24) [Computer program]. Retrieved June 19, 2006, from http://www.praat.org/

Borsky, S., Shapiro, L. P., & Tuller, B. (2000). The temporal unfolding of local acoustic information and sentence context. *Journal of Psycholinguistic Research, 29,* 155–168.

Cooper, A. M. (1991). Laryngeal and oral gestures in English /p, t, k/. In *Proceedings of the XIIth International Congress of Phonetic Sciences, Aix-en-Provence* (Vol. 2, pp. 50–53).

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language, 31,* 218–236.

Cutler, A., & Norris, D. G. (1988). The role of stressed syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 113–121.

Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 498–513.

Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 218–244.

Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of vowel-initial syllables as a function of prosodic structure. *Journal of Phonetics, 24,* 423–444.

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America, 101,* 3728–3740.

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.

Giovanardi, M., & Di Benedetto, G.-B. (1998). Acoustic analysis of singleton and geminate fricatives in Italian. *European Journal of Language and Speech, 1,* 1–13.

Gow, D. W., & Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 344–359.

Haarmann, H. J., & Kolk, H. H. J. (1994). On-line sensitivity to subject–verb agreement violations in Broca's aphasics: The role of syntactic complexity and time. *Brain and Language, 45,* 493–516.

Hughes, A., & Trudgill, P. (1996). *English accents and dialects: An introduction to social and regional varieties of English in the British Isles.* London: Arnold.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language, 44,* 548–567.

Kilborn, K., & Moss, H. (1996). Word monitoring. *Language and Cognitive Processes, 11,* 689–694.

Ladefoged, P. (2001). *A course in phonetics* (4th ed.). Fort Worth, TX: Harcourt Brace.

MacDonald, M. C. (1993). The interaction of lexical and syntactic ambiguity. *Journal of Memory and Language, 32,* 692–715.

Marslen-Wilson, W. D., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition, 8,* 1–71.

Mattys, S. L. (2004). Stress versus coarticulation: Toward an integrated approach to explicit speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 397–408.

Mattys, S. L. & Melhorn, J. F. (in press). Sentential, lexical, and acoustic effects on the perception of word boundaries. *Journal of the Acoustical Society of America.*

Mattys, S. L., Pleydell-Pearce, C. W., Melhorn, J. F., & Whitecross, S. E. (2005). Detecting silent pauses in speech: A new tool for measuring on-line lexical and semantic processing. *Psychological Science, 16,* 958–964.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General, 134,* 477–500.

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

McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language, 39,* 21–46.

Meng, X. L., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin, 111,* 172–175.

Millotte, S., René, A., Wales, R., & Christophe, A. (2007). *Phonological phrase boundaries constrain the on-line syntactic analysis of spoken sentences.* Manuscript submitted for publication.

Nakatani, L. H., & Dukes, K. D. (1977). Locus of segmental cues to word juncture. *Journal of the Acoustical Society of America, 62,* 714–719.

Nakatani, L. H., & Schaffer, J. A. (1978). Hearing "words" without words: Prosodic cues for word perception. *Journal of the Acoustical Society of America, 63,* 234–245.

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

Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1209–1228.

Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology, 34,* 191–243.

Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America, 54,* 1235–1247.

Quené, H. (1992). Durational cues for word segmentation in Dutch. *Journal of Phonetics, 20,* 331–350.

Quené, H. (1993). Segment durations and accent as cues to word segmentation in Dutch. *Journal of the Acoustical Society of America, 94,* 2027–2035.

Raghunathan, T. E., Rosenthal, R., & Rubin, D. B. (1996). Comparing correlated but nonoverlapping correlations. *Psychological Methods, 1,* 178–183.

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition, 90,* 51–89.

Sanders, L. D., & Neville, H. J. (2000). Lexical, syntactic, and stress-pattern cues for speech segmentation. *Journal of Speech, Language, and Hearing Research, 43,* 1301–1321.

Sanders, L. D., & Neville, H. J. (2003). An ERP study of continuous speech processing: I. Segmentation, semantics, and syntax in native speakers. *Cognitive Brain Research, 15,* 228–240.

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

Tabossi, P., Burani, C., & Scott, D. R. (1995). Word recognition in connected speech in Italian. *Journal of Memory and Language, 34,* 440–467.

Turk, A. E., & Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics, 28,* 397–440.

Tyler, L. K., & Marslen-Wilson, W. D. (1986). The effects of context on the recognition of polymorphemic words. *Journal of Memory and Language, 25,* 741–752.

Tyler, L. K., & Warren, P. (1987). Local and global structure in spoken language comprehension. *Journal of Memory and Language, 26,* 638–657.

Tyler, L. K., & Wessels, J. (1983). Quantifying contextual contributions to word-recognition processes. *Perception & Psychophysics, 34,* 409–420.

Van Petten, C., & Kutas, M. (1991). Influences of semantic and syntactic context on open- and closed-class words. *Memory & Cognition, 19,* 95–112.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language, 40,* 374–408.

White, L. S. (2002). *English speech timing: A domain and locus approach.* Unpublished doctoral dissertation, Edinburgh University, Edinburgh, Scotland, United Kingdom.

Appendix A

Stimuli in Experiment 1

Verb–object phrase	Case	Subject phrase
Takes pins/take spins	S	That woman
	P	Those women
Loves parks/love sparks	S	The child
	P	The children
Eats eels/eat seals	S	The shark
	P	The sharks
Starts miles/start smiles	S	The laugh
	P	All the laughs
Counts pears/count spares	S	The worker
	P	The workers
Cuts lips/cut slips	S	The knife
	P	The knives
Cracks nails/crack snails	S	One slight knock
	P	Many slight knocks
Helps pain/help Spain	S	The vital agent
	P	Those vital agents
Nears top/near stop	S	A red train
	P	Some red trains
Makes pace/make space	S	Our leader
	P	Our leaders
Prints talk/print stalk	S	The new typist
	P	The new typists
Likes ports/like sports	S	The sailor
	P	Those sailors
Employs traps/employ straps	S	One novice
	P	Two novices
Dislikes lumps/dislike slumps	S	The medic
	P	Some medics
Takes train/take strain	S	The weightlifter
	P	Those weightlifters

Note. S = singular; P = plural.

Appendix B

Acoustic Analyses for the Test Phrases of Experiments 1 and 2

Following quantitative and qualitative analyses of the stimuli using the Praat speech editor (Boersma & Weenink, 2006), we established that the word boundaries in the "take#spins" and "takes#pins" phrases were indexed by a number of allophonic variations. When the object noun began with a voiceless stop in the singular condition (e.g., "takes pins"), there was generally much greater aspiration of the stop than there was in the plural condition (e.g., "take spins"), in which the stop was preceded by the fricative /s/. There was, in the plural condition, occasionally audible coarticulation between the /s/ and the following continuant (e.g., the /l/ was audible in the preceding frication in "cut#slips," but not in "cuts#lips"). When the object noun began with a vowel in the singular condition (e.g., "eats#eels"), there was glottalization at the initiation of vocalization (a known correlate of word and higher level juncture; e.g., Dilley, Shattuck-Hufnagel, & Ostendorf, 1996). This glottalization was absent in the singular condition, in which the vowel was not word-initial (e.g., "eat#seals"). In most cases, the central /s/ was realized as voiceless, even when it was preceded in the same syllable by a voiced segment (e.g., "nears#top" vs. "near#stop"). Rarely, there was a realizational difference in the central /s/. However, in "employs#traps," the final segment of the first word was realized as /z/ as opposed to the word-initial /s/ in "employ#straps."

Durational analyses on the body and the coda of the target words revealed no reliable difference between acoustic conditions or between recording styles. The duration of the entire phrase, however, was found to be longer in the natural-allophony (1,279 ms) than the concatenated (1,089 ms) conditions, $F(1, 56) = 15.87$, $p = .001$, but it did not interact with the

Table B1
Segment Durations (in Milliseconds) According to Lexical Affiliation of the Central /s/

Segment and recording	take#spins	takes#pins
Central /s/		
Decoarticulated	131	173
Natural allophony	163	149
Following consonant		
Decoarticulated	85	79
Natural allophony	128	201

Note. # = an acoustic juncture.

(*Appendixes continue*)

acoustic condition. However, a key acoustic correlate of the lexical affiliation of central /s/ (or, rarely, /z/) was its duration, as shown in Table B1. This varied according to the recording conditions. In the decoarticulation condition, the central /s/ was longer word-finally than word-initially. In the natural-allophony condition, it was longer word-initially than word-finally. An analysis of variance with recording and acoustics as factors showed no main effects, but a significant Recording × Acoustics interaction, $F(1, 56) = 10.03$, $p < .005$. The difference in /s/ duration was significant in the decoarticulated condition, $t(28) = 3.48$, $p < .005$, but not in the natural condition, $t(28) = 1.08$, $p = .29$.

The duration of the following consonant (e.g., the /p/ in "take#spins" vs. "takes#pins") also varied according to lexical affiliation and recording conditions. In the decoarticulated condition, this consonant was shorter word-initially. In the natural-allophony condition, it was longer word-initially. An analysis of variance with recording and acoustics as factors showed main effects of acoustics, $F(1, 52) = 13.40$, $p = .001$, and recording, $F(1, 52) = 79.08$, $p < .001$. The latter effect was probably attributable, at least in part, to the slightly slower speech rate in the natural-allophony condition as well as the decoarticulation editing procedure (discussed below). There was no significant difference in the duration of the consonant following the central /s/ in the decoarticulated condition, $t(26) = 0.70$, $p = .49$, but the difference was significant in the natural condition, $t(26) = 4.36$, $p < .001$.

These durational analyses suggest that there were two distinct patterns of variation of the word-initial consonant, according to the recording conditions. Word-initial lengthening (e.g., /s/ in "take#spins," /p/ in "takes#pins") was evidenced in the natural-allophony condition. This is in agreement with previous research on juncture-related durational variation (Cooper, 1991; Oller, 1973; Turk & Shattuck-Hufnagel, 2000): Onset consonants in the initial syllables of utterance-medial words are longer than onset consonants in word-medial positions, and word-initial onsets may be lengthened further at higher level prosodic boundaries (Fougeron & Keating, 1997). In the decoarticulation condition, the durational pattern was reversed, with word-initial shortening. This indicates that decoarticulation was realized as a full-utterance boundary. Indeed, at the initial edge of an utterance, most classes of onset consonants appear to be shorter than word-initial onset consonants away from the utterance edge (Fougeron & Keating, 1997; White, 2002). In the case of word-initial stops, this shortening was reinforced by the editing technique used in the decoarticulation condition, which tended to reduce the closure duration of stop consonants following the point of decoarticulation.

Despite the differences between recording conditions in the durational correlates of the boundary, the results of the perceptual studies clearly suggest that both significant lengthening and significant shortening of the onset can—in conjunction with the processes of allophony described above—reinforce the percept of a preceding word boundary.

Appendix C

Main Statistical Analyses With an Alternative Locus for Latency Measurement

Although somewhat arbitrary and subject to idiosyncrasies across conditions, latency analyses performed from target onset revealed similar response patterns. We measured latencies to the long targets (e.g., "spins") from the onset of the /s/ frication. The onset of the short targets (e.g., "pins") was arbitrarily measured as the offset of the frication of the /s/ sound preceding the target (e.g., the /s/ preceding "pins"). This post-/s/ locus might not accurately reflect the actual onset of the target, but given the wide range of target onsets as well as substantial acoustic differences between the concatenation and natural-allophony conditions, this solution appeared to be the most objective way of proceeding—although not as objective and systematic as anchoring latencies to the middle of the stable part of the target nucleus, as we did in the main analyses. Although anchoring the latencies to target onset resulted in a substantial target effect—with longer latencies for long than for short targets (e.g., Experiment 1: 748 ms vs. 592 ms; Experiment 2: 661 ms vs. 531 ms; Experiment 3B: 743 ms vs. 554 ms), attributable to the quantity of information to process before identification—the key findings held. These are succinctly reported below for each experiment.

Experiment 1

An Acoustics × Target interaction, $F_1(1, 88) = 23.07$, $p < .001$; $F_2(1, 14) = 3.57$, $p = .08$, modulated by context, $F_1(2, 176) = 5.39$, $p < .01$; $F_2(2, 28) = 1.60$, $p = .22$, suggested that listeners relied on the acoustic cues differently across the three syntactic contexts, at least by subjects. Although all three contexts showed an Acoustics × Target interaction (i.e., detection latencies were shorter when targets aligned with their respective acoustic cues), this interaction was less pronounced in the singular than in the neutral conditions, $F_1(1, 88) = 7.43$, $p < .01$; $F_2(1, 14) = 12.02$, $p < .005$, and less pronounced in the plural than in the neutral conditions, $F_1(1, 88) = 9.22$, $p < .005$; $F_2(1, 14) = 6.89$, $p < .05$. This interaction did not differ between singular and plural, $F_1(1, 88) < 1$; $F_2(1, 14) < 1$. Thus, although reliance on the acoustic cues was not completely eliminated when syntactic information was available, it was strongly attenuated, as noted in the main analyses. The difference between the singular and plural conditions observed in the main analyses emerged clearly in the target analyses. Despite the generally shorter latencies to long than to short targets, $F_1(1, 88) = 255.82$, $p < .001$; $F_2(1, 14) = 208.15$, $p < .001$, the target effect was modulated by context, $F_1(2, 176) = 7.77$, $p = .001$; $F_2(2, 28) = 3.35$, $p < .005$. The target effect was comparable in the neutral and singular conditions, $F_1(1, 86) < 1$; $F_2(1, 14) < 1$, but both effects differed from that in the plural condition, $F_1(1, 86) = 20.03$, $p < .001$; $F_2(1, 14) = 5.36$, $p < .05$, and $F_1(1, 86) = 28.26$, $p < .001$; $F_2(1, 14) = 5.14$, $p < .05$, respectively. Thus, as in the main analyses, providing a singular context did not result in an advantage for the singular-consistent target relative to the neutral condition, whereas providing a plural context did cause shorter latencies for the plural-consistent target relative to the neutral condition. None of these patterns were significantly affected by recording in subjects and items analyses simultaneously.

Experiment 2

As in Experiment 1, an Acoustics × Target interaction, $F_1(1, 88) = 26.15$, $p < .001$; $F_2(1, 14) = 6.61$, $p < .05$, modulated by context, $F_1(2, 176) = 3.98$, $p < .05$; $F_2(2, 28) = 1.66$, $p = .22$, suggested that listeners relied on the acoustic cues differently across the three syntactic contexts, at least by subjects. The Acoustics × Target interaction, which highlights listeners' use of acoustic cues for segmentation, was less pronounced in the plural than in the neutral condition, $F_1(1, 88) = 7.17$, $p < .01$; $F_2(1, 14) = 3.50$, $p = .08$, and less pronounced in the plural than in the singular condition, $F_1(1, 88) = 4.86$, $p < .05$; $F_2(1, 14) = 1.66$, $p = .22$, although these patterns did not emerge as strongly in the items analyses. The Acoustics × Target interaction did not differ in the neutral and singular conditions, $F_1(1, 88) < 1$; $F_2(1, 14) < 1$. Thus, listeners' reliance on acoustic cues was equally strong when no sentential context or a singular context was available. It was attenuated when a plural context was present. As for the effect of target, once again, despite the generally shorter latencies to the long than to short targets, $F_1(1, 88) = 288.12$, $p < .001$; $F_2(1, 14) = 53.88$, $p < .001$, the target effect was modulated by context, $F_1(2, 176) = 17.05$, $p < .001$; $F_2(2, 28) = 7.77$, $p = .01$. Although comparable in the neutral and singular conditions, $F_1(1, 86) < 1$; $F_2(1, 14) < 1$, the target effect in the plural condition differed from that in both the neutral condition, $F_1(1, 86) = 38.45$, $p < .001$; $F_2(1, 14) = 15.41$, $p < .005$, and the singular condition, $F_1(1, 86) = 22.99$, $p < .001$; $F_2(1, 14) = 7.76$, $p = .01$. Thus, as in the main analyses, and relative to the neutral condition, a singular context did not benefit the singular-consistent targets, whereas a plural context caused shorter latencies for the plural-consistent targets. None of these patterns were significantly affected by recording in subjects and items analyses simultaneously.

Experiment 3B

With latencies measured form target onset, the analysis of variance in this experiment showed a substantial target effect attributable to the new anchor point, with shorter latencies to long than to short targets, $F_1(1, 88) = 308.29$, $p < .001$; $F_2(1, 14) = 43.21$, $p < .001$. More critically, though, the analysis also showed the expected Context × Target interaction, $F_1(2, 176) = 25.44$, $p < .001$; $F_2(2, 28) = 19.12$, $p < .001$. Relative to the neutral condition, the target effect showed an advantage for the singular-consistent targets in the singular condition, $F_1(1, 88) = 7.00$, $p = .01$; $F_2(2, 28) = 5.32$, $p < .05$, and an advantage for the plural-consistent targets in the plural condition, $F_1(2, 176) = 21.54$, $p < .001$; $F_2(2, 28) = 15.25$, $p < .005$. None of these effects were significantly affected by the group factor (i.e., whether the sentential context was supplemented by semantic information). To compare the magnitude of the syntactic effect in the singular and plural conditions, we measured the size of the target effect in those two conditions once the target effect in the neutral condition was subtracted. An analysis of variance with group (syntax only, syntax–semantics) and context (singular, plural) did not show a group effect, $F_1(1, 88) = 2.01$, $p = .16$; $F_2(1, 14) = 1.28$, $p = .28$, a context effect, $F_1(1, 88) < 1$; $F_2(1, 14)$ 1.41, $p = .25$, or a Group × Context interaction, $F_1(1, 88) < 1$; $F_2(1, 14) < 1$. Therefore, as in the main analyses, the magnitude of the syntactic effect was comparable in the singular and plural conditions, and there was no evidence that the added semantics benefited segmentation over and above the advantage provided by syntax.

Appendix D

Stimuli in Experiment 2

Verb–object phrase	Case	Subject phrase
Takes pins/take spins	S	The new machine at the bowling alley
	P	Boys like it when their go-karts
Loves parks/love sparks	S	The energetic child
	P	On bonfire night, all children
Eats eels/eat seals	S	The seafood lover
	P	Some Eskimo tribes
Starts miles/start smiles	S	Beginning the marathon, the sound of the pistol
	P	At the comedy festival, the jokes quickly
Counts pears/count spares	S	The fruit picker
	P	Those junior mechanics
Cuts lips/cut slips	S	The explorer says that the biting arctic wind
	P	To create the dress, a team of seamstresses
Cracks nails/crack snails	S	She tries to avoid doing DIY as it
	P	When walking, the Buddhists are careful not to
Helps pain/help Spain	S	A drugs company has produced a revolutionary Paracetamol that
	P	As part of the treaty, England, France and Italy have agreed to
Nears top/near stop	S	The mountaineer is likely to feel renewed energy as he
	P	All the cars have to slow down to the point where they
Makes pace/make space	S	The race favorite has gone into the lead and now
	P	They've rearranged the furniture in the office to
Prints talk/print stalk	S	The new gossip columnist
	P	Looking for typos in the new gardening book the editors were sure to
Likes ports/like sports	S	After being at sea for long periods of time, the sailor
	P	They love to play football, tennis, polo and in general
Employs traps/employ straps	S	To cut down the number of pheasants eaten by foxes, the gamekeeper
	P	To avoid accidents, the bungee-jump workers are careful when they
Dislikes lumps/dislike slumps	S	The cancer medic
	P	Those governments' economic ministers
Takes train/take strain	S	Because the roads are so congested, the commuter
	P	At the beginning of the tug of war, the referee tells the teams to

Note. S = singular; P = plural.

Received October 20, 2005
Revision received November 1, 2006
Accepted November 6, 2006 ■