

Anticipating Upcoming Words in Discourse: Evidence From ERPs and Reading Times

Jos J. A. Van Berkum

University of Amsterdam and the F. C. Donders Centre for
Cognitive Neuroimaging

Colin M. Brown

Max Planck Institute for Psycholinguistics

Pienie Zwitserlood

Westfälische Wilhelms-Universität Münster

Valesca Kooijman and Peter Hagoort

Max Planck Institute for Psycholinguistics and F. C. Donders
Centre for Cognitive Neuroimaging

The authors examined whether people can use their knowledge of the wider discourse rapidly enough to anticipate specific upcoming words as a sentence is unfolding. In an event-related brain potential (ERP) experiment, subjects heard Dutch stories that supported the prediction of a specific noun. To probe whether this noun was anticipated at a preceding indefinite article, stories were continued with a gender-marked adjective whose suffix mismatched the upcoming noun's syntactic gender. Prediction-inconsistent adjectives elicited a differential ERP effect, which disappeared in a no-discourse control experiment. Furthermore, in self-paced reading, prediction-inconsistent adjectives slowed readers down before the noun. These findings suggest that people can indeed predict upcoming words in fluent discourse and, moreover, that these predicted words can immediately begin to participate in incremental parsing operations.

Keywords: discourse context, lexical anticipation, prediction-sensitive parsing, grammatical gender, EEG

If we did not have the capacity to anticipate, most of us would probably be dead. Anticipation is at the heart of survival. It prevents most of us from keeping poisonous snakes as pets and from going out into a blizzard without a coat. It allows us to predict that we can find dinner in the local supermarket and need money

to pay for it. Anticipation helps us cross the street, catch a frisbee in our hand instead of in our face, and select a mate with whom we stand a chance at reproduction.

With anticipation being important for us humans in so many domains of our lives, it is not unreasonable to expect anticipatory behavior in our use of language as well. And indeed, there is evidence for such behavior. For instance, we routinely predict our upcoming turns in conversation from a variety of subtle cues, including pitch and durational aspects of our interlocutor's current utterance (e.g., Sachs, Schegloff, & Jefferson, 1974; Wennerstrom & Siegel, 2003). At the other end of the spectrum, one might say, is the rather simple anticipation afforded by word-word associative and semantic priming (e.g., Meyer & Schvaneveldt, 1971). And, somewhere between conversational turn taking and intralexical priming is the syntactic garden path phenomenon (e.g., Mitchell, 1994), which can be taken to reflect anticipation of a syntactic structure that once looked promising but turned out to be a dead end.

In this study, we investigated whether listeners and readers can exploit their knowledge of the wider discourse—the linguistic exchange that precedes the currently unfolding sentence—to routinely anticipate specific upcoming words. So, by the time people have arrived at, say, the final determiner in Example 1, are they by any chance expecting any specific word as a plausible continuation?

The burglar had no trouble locating the secret family safe. Of course, it was situated behind a . . . (1)

Various phenomena suggest that listeners and readers might indeed be able to predict specific upcoming words. One is that in

Jos J. A. Van Berkum, Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands and F. C. Donders Centre for Cognitive Neuroimaging, Nijmegen, The Netherlands; Colin M. Brown, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands; Pienie Zwitserlood, Psychologisches Institut II, Westfälische Wilhelms-Universität Münster, Münster, Germany; Valesca Kooijman and Peter Hagoort, F. C. Donders Centre for Cognitive Neuroimaging and Max Planck Institute for Psycholinguistics.

Experiment 1 was conducted at the Max Planck Institute for Psycholinguistics, Experiment 2 was conducted at the F. C. Donders Centre for Cognitive Neuroimaging, and Experiment 3 was conducted at the University of Amsterdam.

We thank René de Bruin, Jesse Jansen, Arnout Koornneef, Marieke van der Linden, Christopher Miller, Bert Molenkamp, Geert-Jan Mertens, Marte Otten, Marcus Spaan, Cathelijne Tesink, Natalia Waaijer, Nienke Weder, and Marlies Wassenaar for their help. This research was supported by a Vernieuwingsimpuls grant from the Netherlands Organization for Scientific Research (NWO) to Jos J. A. Van Berkum, a grant from the Deutsche Forschungsgemeinschaft to Pienie Zwitserlood and Peter Hagoort, and by NWO Grant 400–56-384 to Colin M. Brown and Peter Hagoort.

Correspondence concerning this article should be addressed to Jos J. A. Van Berkum, University of Amsterdam, Department of Psychology 15, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands. E-mail: berkum@psy.uva.nl or j.j.a.vanberkum@uva.nl

natural conversation, interlocutors can “take over” and finish each other’s sentences quite successfully (also noted by Pickering & Garrod, 2004). Furthermore, when people stutter, listeners often seem to have the feeling that they know what they want to say. Finally, when readers are asked to complete a truncated story like Example 1 in a so-called story completion or *cloze* test, many of them come up with the same word (in this case, *painting*). All of this suggests that in at least some circumstances, people can indeed use their knowledge of the wider discourse to predict specific upcoming words. Of course, one might object that comprehenders may be able to do this only when given ample time, for example, because their conversational partners hesitate or, in the paper-and-pencil cloze test, because the utterance simply stops unfolding. The issue we examined, therefore, was whether people can use their knowledge of the wider discourse rapidly enough to predict specific upcoming words “on the fly,” as the current sentence is unfolding.

Does Context-Based Word Prediction Make Sense?

The idea that people might routinely anticipate or predict specific linguistic content in a way that goes beyond a simple intralexical priming mechanism has never been a very popular one in psycholinguistics. With the notable exception of Altmann’s (1997) *The Ascent of Babel*, the authors of recent psycholinguistics textbooks (e.g., Harley, 2001; Jay, 2003; Whitney, 1998) make no reference to the possibility that people might predict upcoming language in this way. Furthermore, prediction has also been notably absent in authoritative monographs and survey chapters on language comprehension (e.g., Cutler & Clifton, 1999; Frazier, 1999; Kintsch, 1998; Perfetti, 1999; Pinker, 1994). The one well-known comprehension model that does have prediction as a fundamental part of its architecture (Elman, 1990; see also Altmann, 1997), although frequently acknowledged as an interesting case of neural network modeling, has been equally lightly discarded as irrelevant to human language comprehension (e.g., see Jackendoff, 2002, p. 59, note 17). Whereas the concept of low-level intralexical priming is ubiquitously accepted as central to understanding human language comprehension, the concept of prediction has instead predominantly acquired a far less favorable association, one with undesirable strategic processing afforded by ill-designed stimuli.

One plausible reason for this state of affairs is that models of language comprehension, in particular those that focus on word recognition, traditionally espouse a strong bottom-up bias. According to classic strictly modular models (Forster, 1979, 1989; cf. Fodor, 1983), for instance, words are recognized solely on the basis of sensory input, and constraining context can only have a *postlexical* impact by affecting the ease with which the word’s syntactic and conceptual properties are integrated with ongoing analyses at syntactic and conceptual levels. However, even more lenient models such as the cohort model (Marslen-Wilson, 1987, 1989), which were regarded as highly interactive at the time of their launching, adhere to a clear bottom-up priority: Sentential and wider context can codetermine the process of selecting (and thus recognizing) the word only after the unfolding word itself has activated a set of lexical candidates. More recent models such as the shortlist model (Norris, 1994) incorporate the very same principle. Of course, fully interactive models that allow for context-

induced lexical *preactivation* or *preselection* have been around for a long time (McClelland & Elman, 1986; McClelland & Rumelhart, 1981; Morton, 1969). However, in the absence of compelling evidence for lexical preactivation, and with several findings that seemed to speak against it either directly (e.g., Connine, 1987, 1990; Samuel, 1981, 1990; Zwitserlood, 1989) or by analogy (e.g., no initial contextual selection of word *sense* either; Swinney, 1979; Tanenhaus, Carlson, & Seidenberg, 1985), few psycholinguists have been inclined to take the idea seriously.

Another important reason for what seems to be a subtle ban on prediction can be found in the enormous impact that generative grammar has exerted on psycholinguistic thinking. Chomsky (1957) and other linguists convincingly argued that language is a *generative system*, allowing the language user to generate an infinite number of expressions from a limited set of elements. The inference that seemed to follow naturally was that, with thousands of linguistic options opening up at every position in an unfolding sentence, it just makes no sense predicting what might come next (see Jackendoff, 2002, p. 59). After all, with speakers allowed to go anywhere they want at just about any time, how could it ever work? And, moreover, what’s the point of telling the future if it is only a few words away and very rapidly recoverable by our highly incremental processing system?

However compelling such arguments might seem, the degree to which listeners and readers make predictions about specific upcoming words is an empirical issue. We agree that there is something slightly odd about conceiving of such predictions within a word recognition perspective, for with no lexical signal having been presented at all, what word is there to recognize? However, a word-recognition perspective is not the only possible view on discourse-based lexical prediction (one reason being that word recognition is not the ultimate goal of the comprehension system; cf. Dahan & Tanenhaus, 2004).

Furthermore, as for the skepticism based on generativity, we note that this makes sense only if the basis for prediction is restricted to syntactic information alone. With syntax being the only predictor, tens of thousands of nouns can indeed follow the determiner *a*, such as *painting*, *wall*, *cloud*, *memo*, *priest*, or *diamond*, and any of these nouns can be preceded by a pronominal modifier, from simple pronominal adjectives like *big*, *invisible*, or *stupid* to complex sentential modifiers like *recently restored but nevertheless still very ugly*. However, language comprehension is more than doing syntax, and speakers usually do not just randomly go wherever syntax allows them to. Semantically speaking, for instance, safes are unlikely to be hidden behind clouds and priests, let alone behind a precious diamond. In addition, speakers tend to adhere to certain conversational maxims (such as the obligation to be relevant, to be clear, and to be specific only when needed; Grice, 1975), which provide strong probabilistic constraints on what the next utterance in a conversation or piece of text might be like (and about). Even sentential phonology can sometimes constrain the options, such as by signaling that the utterance is about to finish or by dictating that in English, the word that immediately follows *a* must begin with a consonant (blocking *a ornament*). Moreover, even though there is perhaps no way of knowing for sure that the very next word is going to be a noun, noun phrases that begin with an indefinite article do tend to have a head noun somewhere, and, with everyday noun phrases, it is bound to come along pretty soon.

Thus, whereas syntax by itself does not provide many cues to the identity of a specific word or to its exact position in the sentence, it can clearly conspire with semantic and other sources of information to converge on a rather plausible specific upcoming word. When native speakers of Dutch were asked to complete a Dutch equivalent of the above example story on paper, some 83% of them used *schilderij* [painting] as the head noun in their completion, in spite of tens of thousands of nominal options afforded by the grammar. We take this convergence, as well as the ability to successfully complete somebody else's sentence, to reflect the language user's talent to very rapidly combine syntactic constraints with the many other sources of information supplied by an unfolding linguistic utterance and its context, and to make intelligent guesses about what might sensibly come next. Whether listeners and readers can do the latter rapidly enough to affect the everyday real-time comprehension of fluently unfolding language, that is, without a momentarily hesitating interlocutor or a patient piece of paper, is the empirical issue on which we now focus.

Prior Research on Context-Based Word Prediction

The question we ask here touches on several well-established research areas. In text comprehension research, for example, considerable effort has been made to determine the extent to which readers make *predictive inferences* from an unfolding piece of text (e.g., Calvo, 2001; Calvo, Meseguer, & Carreiras, 2001; Campion & Rossi, 2001; Fincher-Kiefer, 1993, 1995, 1996; Graesser, Singer, & Trabasso, 1994; Keefe & McDaniel, 1993; Klin, Guzman, & Levine, 1999; Linderholm, 2002; McKoon & Ratcliff, 1992; Murray & Burke, 2003; Murray, Klin, & Myers, 1993; Schmalhofer, McDaniel, & Keefe, 2002; Weingartner, Guzmán, Levine, & Klin, 2003; Whitney, Ritchie, & Crane, 1992). The emerging consensus is that readers do not rigidly infer everything logically possible all the time but do make predictive inferences under particular circumstances, such as when the text is sufficiently constraining and world knowledge makes the inference sufficiently available (e.g., Weingartner et al., 2003). However, predictive inference research has invariably focused on whether readers spontaneously anticipate certain conceptual developments in the unfolding narrative and augment their situation models accordingly, for example, whether reading about an angry husband having thrown a fragile porcelain vase against the wall prompts the reader to infer that the vase probably broke as the result of that. Such a conceptual prediction about the world modeled in one's situation model can of course lead the comprehender to also make a prediction about impending linguistic communication. However, this is by no means a necessity. In this study, we specifically examined whether readers or listeners can exploit their situation model (as well as, presumably, their knowledge about language, communication, the speaker, and the world) to predict specific upcoming words in an unfolding utterance, such as the word *broken* after *He was sorry the vase had . . .*

As mentioned before, models of word recognition that embody the principle of bottom-up priority (e.g., Forster, 1979, 1989; Marslen-Wilson, 1987, 1989; Norris, 1994; see also Cutler & Clifton, 1999) take a clear stance against prediction as being relevant to word recognition, and a number of spoken word recognition studies can be taken to support this position (Connine, 1987, 1990; Grosjean, 1980; Marslen-Wilson & Tyler, 1980; Sam-

uel, 1981, 1990; Zwitserlood, 1989; see Zwitserlood, 1998, for an overview). It is interesting to note, however, that a closer look at the language materials used in testing for word *preactivation* reveals that in many spoken-language studies, the level of contextual constraint was actually quite moderate, with cloze values of 20%–30% being quite common. Although this may have been sufficient to locate the impact of context within the access–selection–integration cascade that is often assumed to subserve word recognition (e.g., Zwitserlood, 1989), it clearly does not provide the strongest possible test for prediction. In the studies presented below, we tested for lexical prediction with spoken ministories that were designed—without resorting to unnatural language use—to be highly predictive at a critical point.

Research on discourse and sentential context effects in written word recognition has uncovered many effects that might be a consequence of context-based lexical prediction. For instance, relative to contextually acceptable but less predictable words, context-predictable words are read more quickly (e.g., Ehrlich & Rayner, 1981; McDonald & Shillcock, 2003; Morris, 1994; Morris & Folk, 1998; Traxler & Foss, 2000; Traxler, Foss, Seely, Kaup, & Morris, 2000; see also Experiment 3 in this article), skipped more often (e.g., Ehrlich & Rayner, 1981; McDonald & Shillcock, 2003; O'Regan, 1979), and responded to more quickly in naming and lexical decision tasks (e.g., Duffy, Henderson, & Morris, 1989; Hess, Foss, & Carroll, 1995; Kleiman, 1980; McClelland & O'Regan, 1981; Schwanenflugel & LaCount, 1988; Schwanenflugel & Shoben, 1985; Schwanenflugel & White, 1991). Unfortunately, however consistent these observations are, they do not make a compelling case for context-based lexical prediction. The reason is that context-induced benefits that are assessed via the predictable word itself can also emerge once the word at hand has been read, because of an easier integration of the associated concept into the wider interpretive context (cf. Foss, 1982; Hess et al., 1995; Traxler & Foss, 2000). Such postlexically facilitated integration may or may not in turn be the consequence of some kind of conceptual anticipation, such as of specific semantic features that might soon become relevant (cf. Federmeier & Kutas, 1999a, 1999b; Schwanenflugel & LaCount, 1988; Schwanenflugel & Shoben, 1985; Van Petten, Coulson, Rubin, Plante, & Parks, 1999). However, even if facilitated integration is the consequence of conceptual anticipation, it does not provide direct evidence for lexical anticipation.

The same ambiguity in interpretation holds for two other empirical phenomena associated with contextual predictability. One is that context-predictable words elicit a smaller N400 in event-related brain potentials (ERPs) than do contextually coherent but less predictable words (e.g., Hagoort & Brown, 1994; Kutas & Hillyard, 1984; Van Petten et al., 1999; see also Experiment 1 in this article). Again, this might reflect the processing benefits of context-based lexical anticipation. However, the extent to which such cloze-dependent N400 effects reflect postlexical facilitated integration, possibly because of conceptual anticipation, but perhaps merely because the story jointly told by context and word is a slightly easier one for which to construct a situation model, is as yet unknown. The second phenomenon, discovered by Federmeier, Kutas, and colleagues, is that anomalous words that are semantically related to context-predictable words elicit smaller N400 effects than do unrelated anomalous words (Federmeier & Kutas, 1999a, 1999b; see also Federmeier, McLennan, De Ochoa, &

Kutas, 2002; Kutas & Federmeier, 2000). In line with earlier behavioral work (Schwanenflugel & Shoben, 1985), this result has been taken as evidence that constraining sentential and wider context can be used to preactivate the lexicosemantic features of the word(s) likely to come next. Under this account, the ERP effect at hand can arise because related anomalous words share some of these features and are as such at a certain processing advantage relative to fully unrelated words. However, the processing advantage of related anomalous words might in principle also emerge from facilitated integration once the word has been presented. To eliminate the latter possibility, Federmeier and Kutas relied on how some of their ERP findings related to off-line plausibility ratings for the items at hand. However, it is obvious that one can obtain a much stronger test for prediction by probing for the selective activation of a particular word before this word or one of its alternatives comes along. In the studies reported below, we probe for the prediction of specific nouns by means of a preceding adjective. Furthermore, we use a word's idiosyncratic and memorized syntactic gender feature to selectively probe for lexical prediction alone.

Experiments 1–3

Our goal of Experiment 1 was to determine whether listeners can use their knowledge of the wider discourse to rapidly predict specific upcoming words as a sentence is unfolding. To examine this, we created a set of predictive two-sentence ministories like *The burglar had no trouble locating the secret family safe. Of course, it was situated behind a . . .*, designed such that, when truncated at the critical indefinite article in a written cloze pretest, the majority of subjects would use the same noun to complete the story (e.g., *painting*). Because the final sentence was always relatively open-ended by itself (*Of course, it was situated behind a . . .*), the predictability of this noun always critically hinged on the wider discourse.

As in German and French, every Dutch noun has a fixed and essentially arbitrary syntactic gender feature, which in indefinite noun phrases (NPs) controls an inflectional suffix on the adjective:

een groot schilderij a big_{neu} painting_{neu} (neuter gender “zero” suffix)
(2)

een grote boekenkast a big_{com} bookcase_{com} (common gender -e suffix)

Because the gender of nouns such as those in Example 2 cannot be derived from their form or meaning, it must be stored with each noun in the mental lexicon (see Van Berkum, 1996, Ch. 2, and references therein; see also Levelt, Roelofs, & Meyer, 1999). In the ERP experiment, we used this fact to probe for discourse-based prediction of a noun before the actual noun itself (or an alternative) was presented. In particular, we first continued the story with an adjective whose inflectional suffix was, in the critical condition, inconsistent with the syntactic gender of the discourse-predictable noun. Subjects were merely asked to listen to the stories as we recorded their electroencephalograms (EEGs). The research logic was simple: If listeners indeed predict a specific noun by the time they have heard the prediction-supporting story up to the indefinite article, an inconsistently gender-inflected adjective should be an unpleasant surprise, and the processing consequences of this perturbation might show up as an ERP effect at the adjective. An

example item is shown in Example 3 below, with the Dutch original followed by an approximate translation in English.

De inbreker had geen enkele moeite de geheime familiekuis te vinden.
[The burglar had no trouble locating the secret family safe.] (3)

Deze bevond zich natuurlijk achter een groot_{neu} maar onopvallend schilderij_{neu}. [Of course, it was situated behind a big-Ø_{neu} but unobtrusive painting_{neu}.] (consistent)

Deze bevond zich natuurlijk achter een grote_{com} maar onopvallende boekenkast_{com}. [Of course, it was situated behind a big-e_{com} but unobtrusive bookcase_{com}.] (inconsistent)

The paradigm we developed here to test for discourse-based lexical prediction before the word itself comes along is actually very similar to the paradigm recently used by Wicha, Kutas, and colleagues (Wicha, Moreno, & Kutas, 2004; see also Wicha, Bates, Moreno, & Kutas, 2003; Wicha, Moreno, & Kutas, 2003). In the most relevant experiment (Wicha et al., 2004), native speakers of Spanish read constraining sentences that were biased toward a particular Spanish noun with a specific syntactic gender (a translated example would be *Little Red Riding Hood carried the food for her grandmother in a . . .*, biased toward *basket_{fem}*). To probe for lexical prediction, Wicha et al. manipulated the gender of the pronominal determiner such that it did or did not agree with the expected noun. The results, which are discussed in more detail in the *General Discussion*, strongly suggest that listeners can use sentential context to predict specific upcoming words.

As can be seen in the item example in Example 3, the stories in our study continued beyond the critical adjective in a fully natural and grammatical way. In stories in which the critical adjective inflection agreed with the discourse-predictable noun, it was this noun (e.g., *painting_{neu}*) that was actually presented. However, in stories in which the critical adjective inflection did not agree with the predictable noun, we avoided overt agreement violations by presenting a semantically coherent alternative noun (e.g., *bookcase_{com}*) that did agree with the prior adjective. Although coherent, these prediction-inconsistent alternative nouns had a much lower discourse-dependent cloze probability than the prediction-consistent nouns they replaced. In isolated sentences, coherent low-cloze words are known to elicit an N400 effect relative to coherent high-cloze words (Hagoort & Brown, 1994; Kutas & Hillyard, 1984; Van Petten et al., 1999). We also know that the N400 is sensitive to discourse-dependent semantic anomalies (Federmeier & Kutas, 1999a, 1999b; St. George, Mannes, & Hoffman, 1994; Van Berkum, Hagoort, & Brown, 1999; Van Berkum, Zwitserlood, Brown, & Hagoort, 2003). These two observations led us to expect a discourse-dependent N400 effect on coherent but prediction-inconsistent nouns like *bookcase* relative to prediction-consistent nouns like *painting*. To prevent this N400 effect from overlapping with the potential ERP effect of a prediction-inconsistent adjective inflection, at least one word separated the critical adjectives from the later noun.

We ran two more experiments to complement this spoken-language EEG study. We conducted Experiment 2, an EEG control study, for reasons explained below. In Experiment 3, we presented a subset of our critical stories for self-paced reading to assess the generality of our earlier findings. The logic was similar to that of Experiment 1: If people predict a specific noun by the time they have processed the prediction-supporting story up to the article, an

incongruently gender-inflected adjective should be an unpleasant surprise that might show up as a reading delay at (or, due to spillover, shortly after) the adjective.

Experiment 1

Method

Subjects. We recruited 24 right-handed native speakers of Dutch (18 women and 6 men, mean age 22 years and range 18–28 years) from the subject pool of the Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands. None had any neurologic impairment, had experienced neurologic trauma, or had used neuroleptics.

Materials. We constructed 74 two-sentence ministories, each of which had a context sentence followed by a critical target sentence; see Example 3 in the introduction to this article. We designed all 74 stories to suggest a specific discourse-predictable noun right after the indefinite article in the target sentence. When given the story up to that point (e.g., *The burglar had no trouble locating the secret family safe. Of course, it was situated behind a . . .*) in a cloze pretest, at least 75% of the 24 respondents in this test spontaneously used the same specific noun to complete the story (e.g., *painting*), with an average cloze probability of 86% ($SD = 6\%$). To make sure that prediction would critically depend on the wider discourse, each target sentence provided little constraint by itself: In a second cloze pretest in which 24 new native speakers of Dutch completed these isolated sentences, an average 6% ($SD = 11\%$) of the respondents came up with the discourse-predictable noun.

In the spoken-language EEG experiment, the indefinite article was always first followed by a gender-inflected critical adjective. In Dutch, adjectives that modify a singular common-gender noun in indefinite noun phrases must have the inflectional suffix *-e*, whereas adjectives that modify a neuter-gender noun have no overtly realized inflectional suffix (i.e., a so-called zero inflection, \emptyset). In prediction-consistent adjectives, the suffix agreed with the grammatical gender of the discourse-predictable noun (e.g., *groot_{neu}* in Example 3), whereas in prediction-inconsistent adjectives (e.g., *grote_{com}* in Example 3), it did not. To avoid confounding our critical manipulation with the actual phonological form of an inflection, the predictable noun was a neuter-gender noun or *het-word* in 34 of the stories (as in Example 3, such that the *-e* inflection is prediction-inconsistent), and a common-gender noun or *de-word* in the remaining 40 stories (such that the \emptyset inflection is prediction-inconsistent), with the two sets of nouns matched on mean predictability.¹ All critical adjectives were semantically acceptable, and because the actual noun was yet to follow, either inflection was grammatically correct at that point in the sentence. The remainder of the target sentence was coherent and grammatical, with the head noun following the critical adjective after at least one intervening other word. In prediction-consistent stories, we used the discourse-predictable noun determined in the story completion pretest (e.g., *painting*), whereas in prediction-inconsistent stories, we used a coherent but much less predictable noun of alternative gender (e.g., *bookcase*). These alternative-gender nouns had an average cloze probability of only 2% ($SD = 3\%$). The Dutch critical items and some sample recordings are available at www.josvanberkum.nl

All stories were recorded with a normal speaking rate and normal intonation by a female native speaker. Each of the two target sentence versions of a particular story was recorded together with the preceding context sentence, with target sentence recording order counterbalanced across condition. A trained native speaker of Dutch identified the acoustic onset of the critical adjective, of the critical inflection therein, and of the critical noun in each target sentence. For each critical adjective, the onset of the inflectional suffix was operationally defined as the point in the acoustic signal where the two versions began to diverge in terms of their respective phonemes. For the *groot-grote* example pair, the stem-final consonant did not differ across versions, and we therefore estimated

inflection onset to be at the onset of the schwa in *grote* and at adjective offset for zero-inflected *groot* (no subsequent word began with a schwa). However, we estimated the inflection onset in pairs like *rood - rode* to be at the onset of the preceding consonant, which was a voiced *d* in *rode* but (due to syllable-final devoicing) an unvoiced *t* in *rood*, as such providing an unambiguous cue to the presence of a zero or schwa inflection, respectively. Across all critical adjectives, and relative to their acoustic onset, mean inflection onset was at 329 ms (range 176–626 ms), and the later noun's mean onset was at 1,039 ms (range 590–1,559 ms). Relative to the onset of the adjective inflection, mean noun onset was on average at 707 ms (range 390–1,290 ms).

In the first of four trial lists, half of the critical set of 74 *de-* and *het-*word stories was presented in prediction-consistent form, and the remainder was presented in prediction-inconsistent form after matching the sets involved on mean cloze value of the discourse-predictable noun in context and in isolation, as well as on mean length (in letters) and sentence position (in words) of the critical adjective. The 74 critical stories and 56 comparable but less predictive stories (cloze between 50% and 75%) were pseudorandomly mixed with 150 spoken filler stories such that no more than 4 critical stories or more than 2 critical stories in either specific condition were immediately consecutive. The filler stories, of which 60 addressed a different issue (see Van Berkum, Brown, Hagoort, & Zwitserlood, 2003, Experiment 2), had an uncontrolled and presumably average level of constraint. Each of five trial blocks began with 2 filler stories. We derived the second list from the first one by rotating the condition of the critical items. We derived two more lists from the first two by reversing the order of these trials. Each list began with 20 practice stories and defined the session for 6 subjects, each of whom never saw an item in more than one condition.

Procedure, EEG recording, and analysis. After electrode application, subjects sat in a sound-attenuating booth and listened to the stimuli over headphones. They were asked to process each story for comprehension. Subjects knew that EEG recording would only occur as they heard the last sentence of each ministory and were asked to avoid eye and other movements during recording. No additional task demands were imposed. After a short practice, the trials were presented in five blocks of 15 min, separated by rest periods.

Each trial consisted of a 300-ms auditory warning tone followed by 700 ms of silence, the spoken discourse context, 1,000 ms of silence, and the spoken final sentence. To inform subjects when to sit still for EEG recording, an asterisk was displayed from 500 ms before onset of the target sentence to 1,000 ms after its offset. The context and target sentences were played from two separate sound files because of a similar constraint imposed on other items presented in the same session (see Van Berkum, Brown, et al., 2003, Experiment 2). The 1,000-ms pause duration between offset of the context sentence and onset of the target sentence was based on the average natural pause between context and target sentences when recorded together, estimated from a representative sample of the materials. An informal pretest as well as later remarks of our EEG subjects indicated that this fixed intersentence pause was experienced as entirely natural, and as such escaped the listeners' attention. Sample sound files with this pause can be downloaded from www.josvanberkum.nl

The EEG was recorded from 29 silver-chloride electrodes, each referred to the left mastoid, in an elastic cap. Five electrodes were placed over the standard 10% system midline sites Fz, FCz, Cz, Pz, and Oz. Nine pairs

¹ The 74 critical items were part of a larger set of 120 items that also included somewhat less predictive stories (cloze value of the discourse-predictable word between 50% and 75%). This larger set contained an equal number of discourse-predictable common- and neuter-gender nouns. We also observed the ERP effects reported for our critical 74 items for that larger gender-balanced set, albeit in somewhat attenuated form (see www.josvanberkum.nl for details).

were placed over the standard lateral sites AF3/AF4, F3/F4, F7/F8, FC3/FC4, FT7/FT8, C3/C4, CP3/CP4, P3/P4, and PO7/PO8. Three additional pairs were placed laterally over symmetrical nonstandard positions: (a) a left (LT) and right (RT) temporal pair placed laterally to Cz at 33% of the interaural distance, (b) a left (LTP) and right (RTP) temporo-parietal pair placed 30% of the interaural distance lateral and 13% of the nasion-inion distance posterior to Cz, and (c) a parietal pair midway between LTP/RTP and PO7/PO8 (LP and RP). Vertical and horizontal eye movements were monitored via a supra- to suborbital bipolar montage and a right-to-left canthal bipolar montage, respectively. We recorded activity over the right mastoid bone to determine whether there were differential contributions of the experimental variables to the left mastoid site (we observed no such differential effects). We amplified the EEG and EOG recordings with a NeuroScan SynAmp Model 5083 EEG amplifier (NeuroScan, Herndon, Virginia), using a hi-cut of 70 Hz and a time constant of 8 s (0.02 Hz). We kept electrode impedances below 3 kOhm for the EEG recording and below 5 kOhm for the EOG recording. The EEG and EOG signals were digitized online at 500 Hz and screened off-line for eye movements, muscle artifacts, electrode drifting, and amplifier blocking in a critical window that ranged from 150 ms before to 2,100 ms after the acoustic onset of the critical adjective (this interval always extended at least 1,000 ms beyond acoustic onset of the later noun). Trials containing such artifacts were rejected (20.5%, with no asymmetry across conditions).

After baseline correcting (by subtraction) the waveforms of the individual trials relative to the relevant (of three) 150-ms prestimulus baseline intervals, we computed average waveforms for each subject and condition relative to the estimated acoustic onset of each of three critical stimulus events: the adjective, the adjective's inflectional suffix, and the later noun. For each of these events, but particularly for the second and third, we screened the ERPs for waveform overlap from preceding events. We observed no such problematic overlap in this study.

In analyses of variance (ANOVAs), we used mean amplitude values computed for each subject in one or more specific latency ranges, defined either on a priori grounds (300–500 ms after noun onset as standard N400 window) or on the basis of the grand-average ERPs (all other latency ranges). We adjusted univariate F tests with more than one degree of freedom in the numerator by means of the Geisser–Greenhouse/Box's epsilon hat correction. We evaluated all results in a midline ANOVA that crossed prediction consistency (consistent or inconsistent with discourse-predictable noun) with a simple five-level midline-electrode factor (Fz, FCz, Cz, Pz, and Oz) and a quadrant ANOVA that fully crossed prediction consistency with hemisphere (left and right) by anteriority (anterior and posterior). The latter analysis effectively defined four quadrants: left-anterior, involving AF3, F3, F7, FC3, and FT7; right-anterior, involving AF4, F4, F8, FC4, and FT8; left-posterior, involving LTP, CP3, LP, P3, and PO7; and right-posterior, involving CP4, RTP, P4, RP, and PO8. If necessary, these two omnibus tests were followed by more specific ANOVAs.

Results

Adjective onset. Figure 1 displays, for each electrode, the grand average event-related brain potentials time-locked to the acoustic onset of the critical adjective for adjectives whose inflection was consistent (solid line) or inconsistent (dotted line) with the gender of the discourse-predictable noun. Also displayed in Figure 1, at Cz, is the range and mean acoustic onset of the critical adjective inflection (i) and of the later noun (n), relative to adjective onset, across the set of items involved.

The most striking effect of inconsistency in Figure 1 is a large negative (upward) deflection emerging around 1,000 ms, right where the prediction-consistent or -inconsistent nouns begin to unfold. As shown later, reaveraging the EEG relative to noun onset

confirms that this late negativity is a noun-elicited discourse-dependent N400 effect.

Figure 1 also shows positive deflections associated with inconsistently inflected adjectives. The largest of these is around 500–800 ms and is most prominent at midline fronto-central sites (e.g., FCz). Mean amplitude ANOVAs in the 500–800-ms latency range revealed no reliable main effect of consistency in the midline and quadrant ANOVAs, $F(1, 23) = 2.31$, $MSE = 9.10$, $p = .142$, and $F(1, 23) = 1.82$, $MSE = 20.10$, $p = .190$, respectively, and no reliable interaction involving this factor, although the midline ANOVA Consistency \times Electrode interaction did approach significance, $F(4, 92) = 2.93$, $MSE = 1.28$, $p = .068$ only. Reliable simple main effects emerged at FCz, FC3, and C3. However, this fronto-central positivity largely overlaps with the latency range for noun onsets, making it difficult to uniquely associate it with the adjective inflection.

An earlier and somewhat more broadly distributed positive deflection around approximately 300–400 ms (extending somewhat beyond the latter at some sites) can also be discerned in Figure 1. Mean amplitude ANOVAs in the 300–400-ms latency range revealed no reliable main effect of consistency in the midline and quadrant ANOVAs, $F(1, 23) = 1.48$, $MSE = 7.58$, $p = .237$, and $F(1, 23) = 2.26$, $MSE = 23.83$, $p = .147$, respectively, and no reliable interaction involving this factor. Also note that the range of measured inflection onsets, schematically indicated below the waveforms measured at Cz, fully overlaps with—and is actually wider than—the latency range in which this small early positivity can be seen. This positive deflection might reflect differential processing of prediction-inconsistent critical inflections, with the associated ERP effect smeared due to inflection onset variability in this adjective onset analysis. If so, we should be able to sharpen and enlarge it if we recompute the waveforms relative to the measured onsets of those inflections.

Adjective inflection onset. Figure 2 displays, for each electrode, the grand average event-related brain potentials time-locked to the acoustic onset of the critical adjective inflection for inflections that were consistent (solid line) or inconsistent (dotted line) with the gender of the discourse-predictable noun. Also displayed in Figure 1, at Cz, is the range and mean acoustic onset of the later noun (n), relative to adjective inflection onset, across the set of items involved.

Again, there is a large late negative deflection in the latency range of the onset of the noun, identified in the *Noun onset* section as a discourse-dependent noun-elicited N400 effect. However, the ERPs computed relative to the critical adjective inflection also reveal a small but clear positive deflection to the inconsistent inflection, emerging somewhere in the first 50 ms after measured inflection onset at all but a few left-posterior electrodes and lasting until about 250 ms after inflection onset at most of those sites. As can be seen in Figure 2 at Cz, the offset of this prediction inconsistency effect was well before the acoustic onset of the later noun, which suggests that it must indeed have been elicited by the inflection.

As can be seen in Table 1, mean amplitude ANOVAs conducted in the 50–250-ms latency range attest to the reliability of this very early positivity. A significant main effect of prediction consistency was obtained in the midline as well as the quadrant ANOVAs. The effect did not significantly vary across midline electrode site, but it did vary across quadrants, with simple main effects revealing a

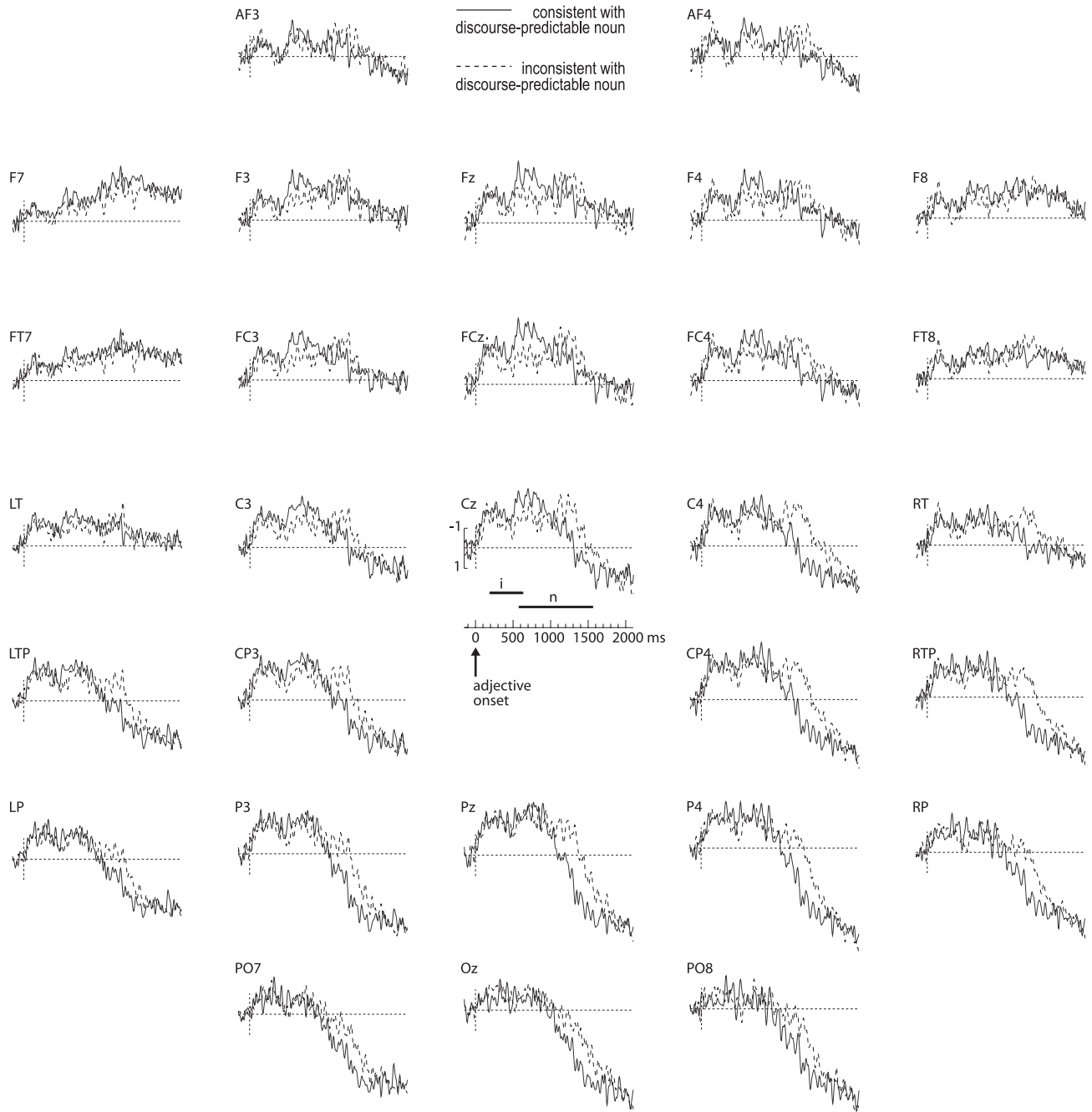


Figure 1. Adjectives in discourse (Experiment 1). Grand average event-related brain potential waveforms time-locked to acoustic onset of the critical adjectives in discourse for adjectives whose inflectional suffix was consistent (solid) or inconsistent (dotted) with the gender of the discourse-predictable noun. In this and all following figures, negative polarity is plotted upward, and horizontal bars at Cz indicate, across all items, the range and mean acoustic onset of the critical inflection (*i*), the later noun (*n*), and sentence end (*e*), relative to 0 ms.

reliable effect in the left-anterior, right-anterior, and right-posterior quadrants (of 0.54 μV , 0.75 μV , and 0.78 μV , respectively). We obtained weaker but comparable prediction consistency effects in supplementary data analyses involving the larger gender-balanced 120-item set, for example, $F(1, 23) = 5.40$, $MSE = 5.88$, $p = .029$, in the omnibus quadrant ANOVA.

At anterior sites, Figure 2 also reveals a second positive deflection around 300–500 ms after inflection onset. However, mean amplitude ANOVAs revealed no reliable prediction consistency main effects in this latency range, for example, $F(1, 23) = 0.74$, $MSE = 11.55$, $p = .398$, and $F(1, 23) = 0.41$, $MSE = 29.95$, $p = .528$ in the midline and quadrant ANOVAs, respectively, and no

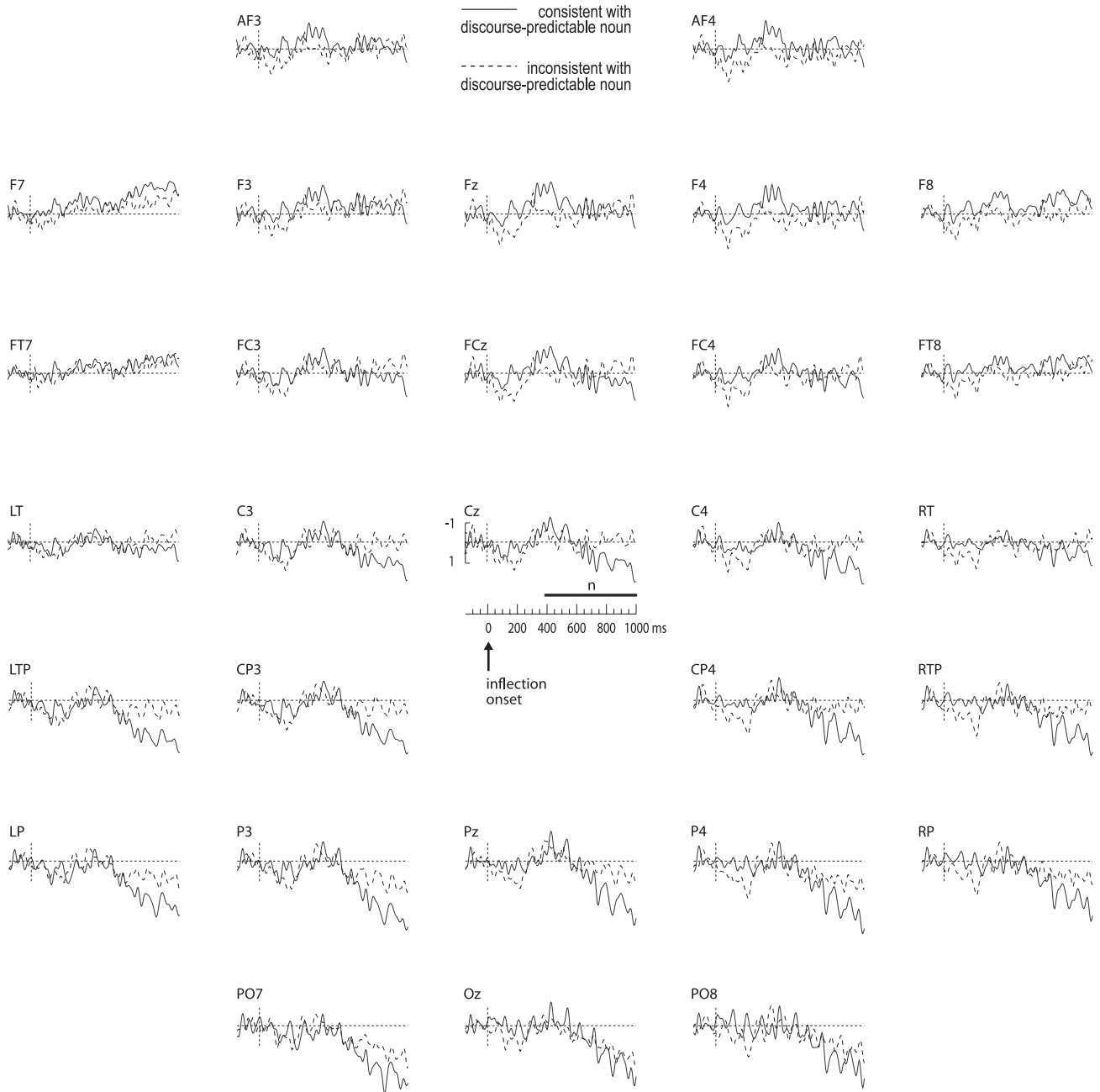


Figure 2. Inflections in discourse (Experiment 1). Grand average event-related brain potential waveforms time-locked to acoustic onset of the critical adjective inflections in discourse for inflectional suffixes that were consistent (solid) or inconsistent (dotted) with the gender of the discourse-predictable noun.

reliable interactions involving (or electrode-specific simple main effects of) this factor.

Noun onset. Figure 3 displays, for each electrode, the grand average event-related brain potentials time-locked to the acoustic onset of the later noun, as a function of whether it was consistent with—that is, identical to—the discourse-predictable noun (solid line) or a prediction-inconsistent alternative noun (dotted line).

As expected, prediction-inconsistent nouns elicited a very sizable N400 effect, which peaked at about 350–400 ms after acous-

tic noun onset (best seen in difference waveforms). The N400 effect is largest at Pz (where it corresponds to a $-2.92 \mu\text{V}$ mean amplitude change in the 300–500 ms latency range), but it can be discerned at all but a single electrode (F7). As might be expected from its size and consistency over electrodes, mean amplitude ANOVAs in the standard N400 latency range of 300–500 ms, displayed in Table 2, confirm that this is a reliable effect.

As can be seen in Figure 3, a sizable differential effect already emerges in the 100–200-ms latency range. In Table 3, we report

Table 1
Analyses of Variance (ANOVAs) on Mean Event-Related Brain Potential Amplitude in the 50–250 Milliseconds After Inflection Onset in Discourse (Experiment 1)

Source amplitude difference (μV)	<i>df</i>	<i>F</i>	<i>MSE</i>	<i>p</i>	
Midline ANOVA (5 electrodes)					
PC	1, 23	5.86	4.98	.024*	
PC \times EI	4, 92	0.36	0.68	.707	
Quadrant ANOVA ($2 \times 2 \times 5$ electrodes)					
PC	1, 23	6.41	13.44	.019*	
PC \times An	1, 23	0.11	5.09	.738	
PC \times He	1, 23	4.08	1.64	.055	
PC \times An \times He	1, 23	4.51	0.20	.045*	
PC \times An \times He \times EI	4, 92	1.35	0.04	.267	
Simple main effects of prediction consistency for each electrode quadrant					
LA	0.54	1, 23	5.16	3.42	.033*
RA	0.75	1, 23	7.51	4.54	.012*
LP	0.32	1, 23	0.94	6.58	.342
RP	0.78	1, 23	6.23	5.83	.020*

Note. For the midline and quadrant ANOVAs, only effects involving prediction consistency are reported: PC = prediction consistency (consistent and inconsistent); EI = electrode; An = anteriority (anterior and posterior); He = hemisphere (left and right). Also shown is the simple main effect of prediction consistency and the associated inconsistent-consistent amplitude difference in μV for each electrode quadrant: LA = left anterior; RA = right anterior; LP = left posterior; RP = right posterior. * $p < .05$.

the results of mean amplitude ANOVAs in this early latency range. Although the waveforms suggest that the early negativity might be distinct from the N400 effect, the two negative deflections have highly comparable scalp distributions. We therefore cannot rule out that the early negativity is simply the ascending flank of a very early N400 effect, with the dip at approximately 200 ms accidentally caused by residual noise (e.g., residual alpha).

Discussion

In the ERPs time-locked to adjective onset, shown in Figure 1, we could discern no clear adjective-elicited effect. However, as can be seen in Figure 2, reaveraging the EEG relative to the acoustic onset of the adjective's inflection uncovered a small but reliable positive deflection in the ERPs elicited by prediction-inconsistent inflections relative to prediction-consistent ones. Because this differential ERP effect hinges on the lexically stored syntactic gender of an expected but not yet presented noun, it suggests that discourse-level information can indeed lead people to anticipate specific upcoming words "on the fly" as a local sentence unfolds. Moreover, because the effect is elicited by an adjective inflection that mismatches the syntactic gender of an upcoming noun but is formally still correct, it also suggests that the syntactic gender properties of a strongly anticipated noun can immediately begin to interact with locally supplied syntactic constraints as part

of a parsing process that takes not only overtly presented but also anticipated structure into account.

These inferences depend solely on the presence of a differential effect (cf. Van Berkum, 2004, in which such sensitivity inferences are contrasted with four other types of inferences that can be supported by ERP data). However, we note that the nature of the present ERP effect, as defined by the combination of polarity, shape, scalp distribution, and coarse timing characteristics, does not straightforwardly remind us of any other ERP effect observed in language comprehension research. We return to this in the *General Discussion* section.

As revealed by their responses in carefully structured postsession interviews, our subjects had not noticed the critical manipulation or the associated critical features of our items. In part, this may be due to the relative salience of certain aspects of the filler items (30 of which contained ambiguous referring expressions; cf. Van Berkum, Brown et al., 2003). However, it also clearly suggests that the generation of strong lexical predictions and the subsequent adjective-based disconfirmation of those predictions does not in itself attract attention. This can be taken to support our hypothesis that such predictions are made routinely and effortlessly, and, in addition, that the selective predictability of our materials (i.e., with high cloze values at certain points in the story only) was sufficiently representative of everyday language to remain unnoticed. It also suggests that the subtle disconfirmation of such predictions (i.e., involving no overt anomaly) is sufficiently normal to escape attention as well. We briefly return to this issue in the *General Discussion* section.

Some important concerns need to be addressed before we can accept and elaborate upon the aforementioned theoretical implications. The most pressing one is that the effect emerges extremely rapidly in the ERP waveforms, somewhere in the first 50–100 ms after measured inflection onset. To rule out the possibility that this effect was an artifact of some uncontrolled accidental difference in acoustic realization across the two sets of critical adjectives, and to simultaneously verify our assumption that the ERP effect critically hinged on information supplied by the prior discourse, we conducted a control EEG experiment in which listeners heard the same critical sentences, played from the same recordings, without the prediction-supporting wider discourse. If the inflection-elicited ERP effect is truly discourse-dependent, it should disappear when the wider discourse is removed. Along the same lines (cf. Van Berkum, Hagoort, & Brown, 1999; Van Berkum, Zwitterlood, et al., 2003), this control study also allowed us to determine the extent to which the sizable N400 effect elicited by coherent but prediction-inconsistent nouns is a discourse-dependent effect.

Experiment 2

Method

Subjects. For Experiment 2, 24 right-handed native speakers of Dutch (18 women and 6 men, mean age 22 years and range 19–29 years) were recruited from the subject pool of the Max Planck Institute for Psycholinguistics. None had any neurologic impairment, had experienced neurologic trauma, or had used neuroleptics. Also, none had participated in Experiment 1.

Materials. In Experiment 2, each subject listened to the same 120 critical target sentences as in Experiment 1 that were now presented without the prediction-supporting wider discourse. In the first of six dif-

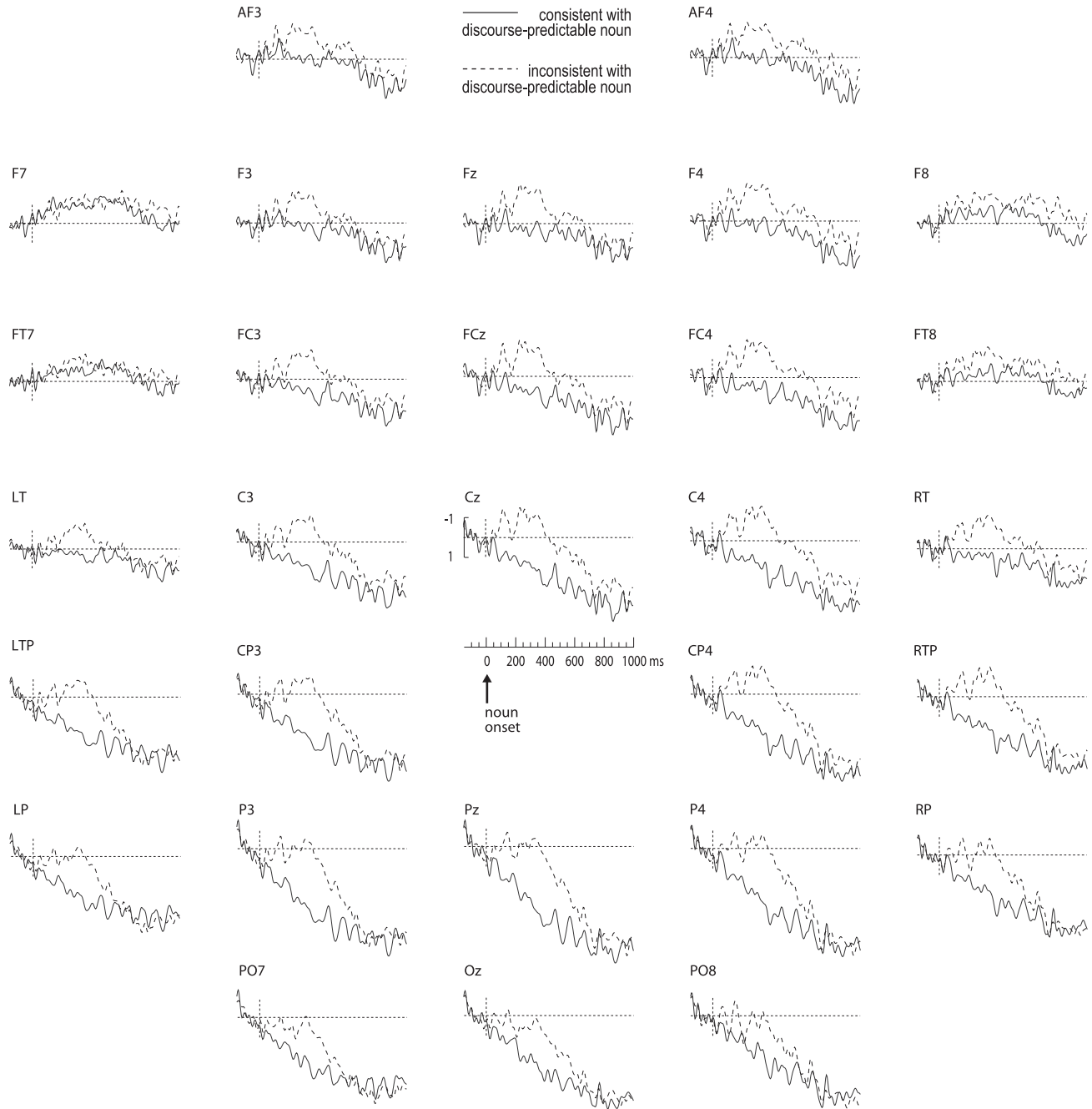


Figure 3. Nouns in discourse (Experiment 1). Grand average event-related brain potential waveforms time-locked to acoustic onset of discourse-predictable nouns (solid) or coherent but less predictable nouns (dotted) in discourse.

ferent trial lists, half of the critical sentences were presented in (formerly) prediction-consistent form, and half were presented in (formerly) prediction-inconsistent form using the same setwise matched item subsets we used for the lists of Experiment 1. The 120 critical sentences were pseudorandomly mixed with 250 filler sentences (180 of which addressed a different issue; see Van Berkum, Zwitserlood, Bastiaansen, Brown, & Hagoort, 2004), such that no more than 4 critical sentences and no more than 2 critical sentences in either consistency condition were immediately consecutive. Each of five 74-sentence trial blocks began with two filler

stories. We derived the second list from the first one by rotating the condition of the critical items while leaving their list position intact. We derived four more lists from the first two by rotating conditions across the 180 presently noncritical sentences while keeping all presently critical items as is. Each list began with 20 practice items and defined the session for 4 subjects, each of whom never saw an item in more than one condition.

Procedure, EEG recording, and analysis. Apart from the materials and some trial timing changes associated with the presentation of a single sentence, the procedure, EEG recording, and analysis were identical to

Table 2
Analyses of Variance (ANOVAs) on Mean Event-Related Brain Potential Amplitude in the 300–500 Milliseconds After Noun Onset in Discourse (Experiment 1)

Source amplitude difference (μV)	<i>df</i>	<i>F</i>	<i>MSE</i>	<i>p</i>	
Midline ANOVA (5 electrodes)					
PC	1, 23	9.67	24.13	.005*	
PC \times EI	4, 92	4.97	0.92	.011*	
Quadrant ANOVA (2 \times 2 \times 5 electrodes)					
PC	1, 23	9.30	62.10	.006*	
PC \times An	1, 23	12.06	8.34	.002*	
PC \times He	1, 23	0.42	7.99	.521	
PC \times An \times He	1, 23	0.25	1.70	.619	
PC \times An \times He \times EI	4, 92	2.71	0.12	.054	
Simple main effects of prediction consistency for each electrode quadrant					
LA	-0.74	1, 23	3.40	9.74	.078
RA	-1.07	1, 23	3.55	19.20	.072
LP	-2.12	1, 23	17.35	15.58	.000*
RP	-2.28	1, 23	8.72	35.61	.007*

Note. For the midline and quadrant ANOVAs, only effects involving prediction consistency are reported: PC = prediction consistency (consistent and inconsistent); EI = electrode; An = anteriority (anterior and posterior); He = hemisphere (left and right). Also shown is the simple main effect of prediction consistency and the associated inconsistent-consistent amplitude difference in μV for each electrode quadrant: LA = left anterior; RA = right anterior; LP = left posterior; RP = right posterior. * $p < .05$.

those of Experiment 1. Each isolated-sentence trial began with a 300-ms warning beep followed after 1,200 ms of silence by a single spoken sentence. To help subjects avoid eye movements, a fixation asterisk was displayed on a computer screen from 1,000 ms before sentence onset to 1,000 ms after sentence offset. The EEG and EOG signals were screened off-line for eye movements, muscle artifacts, electrode drifting, and amplifier blocking in a critical window that ranged from 150 ms before to 1,200 ms after acoustic onset of the critical adjective inflection, and in the equivalent window time-locked to acoustic onset of the noun. Trials containing such artifacts were rejected (9.7%, with no condition asymmetry).

Results

Adjective inflection onset. Figure 4 displays, for each electrode, the grand average event-related brain potentials time-locked to the acoustic onset of the critical adjective inflection for inflections that were consistent (solid line) or inconsistent (dotted line) with the gender of the formerly discourse-predictable noun. Also displayed in Figure 4, at Cz, is the range and mean acoustic onset of the later noun (*n*), relative to adjective inflection onset, across the set of items involved.

Whereas Figure 2 showed that critical prediction-inconsistent inflections elicited a distinct and widely distributed positive deflection, Figure 4 reveals that the very same critical inflections do not elicit a reliable effect if the prediction-supporting discourse is taken away. Although a small negative trend emerges in the

relevant 50–250-ms latency range at several sites, the associated mean amplitude statistics displayed in Table 4 provide no evidence for a reliable differential effect.

Noun onset. Figure 5 displays, for each electrode, the grand average event-related potentials time-locked to the acoustic onset of the noun as a function of whether this noun had in Experiment 1 been the discourse-predictable noun (solid line) or its prediction-inconsistent alternative (dotted line).

As expected, and as confirmed by the 300–500-ms mean amplitude ANOVA results displayed in Table 5, the substantial discourse-dependent N400 effect that was obtained with these nouns when they were embedded in a prediction-supporting discourse context in Experiment 1 (see Figure 3) was not observed when the wider discourse was removed.

The waveforms in Figure 5 actually do begin to diverge after some 500 ms from noun onset in a latency range that is not associated with the standard sentence- and discourse-dependent N400 effect. Mean amplitude ANOVAs in the 500–700-ms window revealed a significant main effect of consistency in the midline analysis, $F(1, 23) = 4.91$, $MSE = 17.77$, $p = .037$, and a related trend in the quadrant analysis, $F(1, 23) = 3.90$, $MSE = 45.89$, $p = .06$, with no significant interactions involving consistency in either analysis.

Table 3
Analyses of Variance (ANOVAs) on Mean Event-Related Brain Potential Amplitude in the 100–200 Milliseconds After Noun Onset in Discourse (Experiment 1)

Source amplitude difference (μV)	<i>df</i>	<i>F</i>	<i>MSE</i>	<i>p</i>	
Midline ANOVA (5 electrodes)					
PC	1, 23	8.16	11.91	.009*	
PC \times EI	4, 92	1.94	0.85	.158	
Quadrant ANOVA (2 \times 2 \times 5 electrodes)					
PC	1, 23	9.62	24.26	.005*	
PC \times An	1, 23	6.90	5.57	.015*	
PC \times He	1, 23	1.97	5.04	.174	
PC \times An \times He	1, 23	2.84	1.17	.105	
PC \times An \times He \times EI	4, 92	3.34	0.12	.053	
Simple main effects of prediction consistency for each electrode quadrant					
LA	-0.26	1, 23	0.70	6.00	.411
RA	-0.91	1, 23	5.10	9.66	.034*
LP	-1.30	1, 23	15.47	6.56	.001*
RP	-1.47	1, 23	9.40	13.82	.005*

Note. For the midline and quadrant ANOVAs, only effects involving prediction consistency are reported: PC = prediction consistency (consistent and inconsistent); EI = electrode; An = anteriority (anterior and posterior); He = hemisphere (left and right). Also shown is the simple main effect of prediction consistency and the associated inconsistent-consistent amplitude difference in μV for each electrode quadrant: LA = left anterior; RA = right anterior; LP = left posterior; RP = right posterior. * $p < .05$.

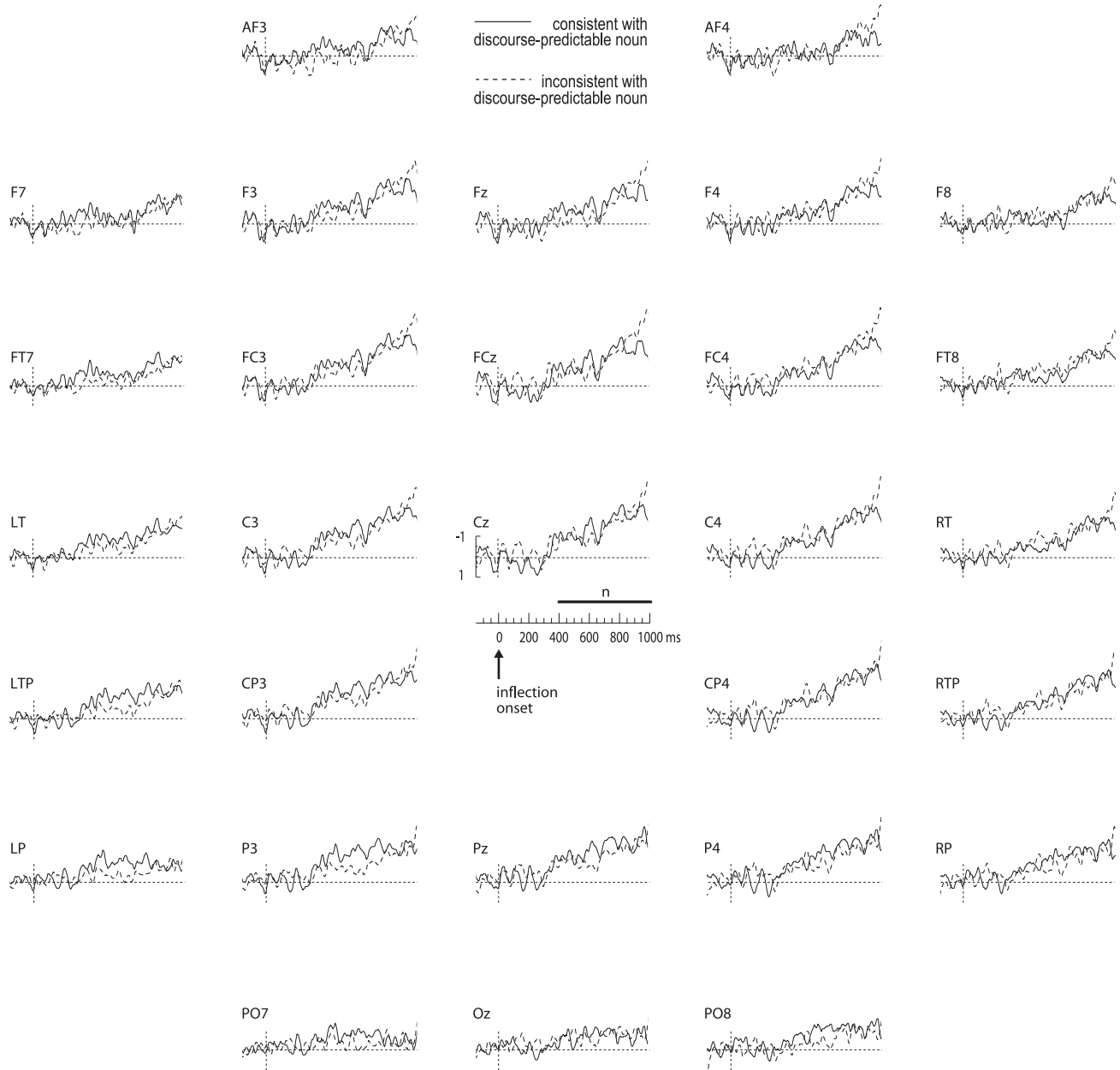


Figure 4. Inflections without prior discourse (Experiment 2). Grand average event-related brain potential waveforms time-locked to acoustic onset of the critical adjective inflections in their local carrier sentence, for inflectional suffixes that were consistent (solid) or inconsistent (dotted) with the gender of the formerly discourse-predictable noun.

Discussion

ERPs at adjective inflections. As revealed by the comparison of Figure 4 with Figure 2, the reliable positive ERP deflection elicited by prediction-inconsistent adjective inflections embedded in a prediction-supporting wider discourse in Experiment 1 was not elicited by the same inflections embedded in an essentially nonpredictive single sentence in Experiment 2. This suggests that the former is no artifact of accidental differences in acoustic realization across the two sets of critical adjectives, but instead

reflects the processing consequences of disconfirming a strong discourse-based lexical prediction.

We were obviously still concerned over the very early onset of the ERP effect. Although statistical analysis did not reveal a significant consistency effect in the 0–50-ms latency range, an examination of the waveforms in Figure 2 does suggest that the effect emerges right at the estimated acoustic onset of the inflection. We know from earlier work (Van Berkum, Zwitserlood, et al., 2003) that discourse-anomalous spoken words can elicit an N400

Table 4
Analyses of Variance (ANOVAs) on Mean Event-Related Brain Potential Amplitude in the 50–250 Milliseconds After Inflection Onset in Nonpredictive Isolated Sentences (Experiment 2)

Source	<i>df</i>	<i>F</i>	<i>MSE</i>	<i>p</i>
Midline ANOVA (5 electrodes)				
PC	1, 23	0.60	7.99	.446
PC × EI	4, 92	0.75	1.39	.466
Quadrant ANOVA (2 × 2 × 5 electrodes)				
PC	1, 23	0.16	19.50	.690
PC × An	1, 23	0.46	4.78	.506
PC × He	1, 23	1.21	2.10	.282
PC × An × He	1, 23	0.50	0.29	.487
PC × An × He × EI	4, 92	1.67	0.15	.199
Simple main effects of prediction consistency for each electrode quadrant				
LA	1, 23	0.11	6.15	.739
RA	1, 23	0.15	8.59	.700
LP	1, 23	0.19	5.45	.665
RP	1, 23	0.77	6.48	.389

Note. For the midline and quadrant ANOVAs, only effects involving prediction consistency are reported: PC = prediction consistency (consistent and inconsistent); EI = electrode; An = anteriority (anterior and posterior); He = hemisphere (left and right). Also shown is the simple main effect of prediction consistency and the associated inconsistent-consistent amplitude difference in μV for each electrode quadrant: LA = left anterior; RA = right anterior; LP = left posterior; RP = right posterior. * $p < .05$.

effect within some 150–200 ms after their acoustic onset, even in so-called “low-constraint stories” in which the anomalous word does not substitute for a strongly expected coherent word (Van Berkum, Zwitserlood, et al., 2003, Figure 3). Thus, we know that the comprehension system can sometimes very rapidly map the unfolding speech signal onto a mental representation of what the wider discourse is about. However, we obviously do not wish to claim that such mapping can occur instantaneously—within the brain, even very simple computations take some tens of milliseconds to unfold.

We believe the explanation for this apparent zero-millisecond delay can be found in details of the procedure we used to determine the acoustic onset of an adjective inflection. As described before, we operationally defined inflection onset as the point in the acoustic signal at which the two adjective variants (e.g., *groot* and *grote*) began to diverge in terms of different phonemes. What we were unable to take into account in this procedure, however, is the fact that the presence or absence of an upcoming inflectional suffix can be signaled by very subtle yet reliable coarticulatory and durational changes in the stem of a word (e.g., Jongman, 1998; Nootboom, 1972) well before the two versions of the adjective diverge in terms of a discretely different phoneme. There is increasing evidence that listeners are in fact very sensitive to these cues (e.g., Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Gaskell & Marslen-Wilson, 2001; Kemps, Ernestus, Schreuder, & Baayen, in press; Kemps, Wurm, Ernestus, Schreuder, & Baayen, 2005; Salverda, Dahan, & McQueen, 2003). Moreover, adding the in-

flexional suffix *-e* to an adjective alters its syllabic structure. We know from other research that Dutch listeners are acutely sensitive to syllable boundary cues present in the speech input (Zwitserlood, 2004). For the adjectives used here, such cues are present as early as the transition between vowel and consonant—which might well be some 100–150 ms earlier than the alignment point used in our EEG analyses. Taken together, there are good reasons to believe that our phoneme-based estimate of the onset of an inflectional suffix is too late, with the critical inflectional information becoming available to our subjects at some unknown earlier moment (possibly even ~100–150 ms before).

ERPs at nouns. As illustrated by the difference between Figures 3 and 5 and confirmed by statistics in the 300–500-ms latency range, the sizable N400 effect that was elicited by prediction-inconsistent nouns (e.g., *bookcase*) relative to their prediction-consistent counterparts (e.g., *painting*) in discourse completely disappeared when the prediction-supporting discourse context was taken away. This suggests, as predicted, and analogous to findings obtained with anomalous spoken words (Van Berkum, Zwitserlood, et al., 2003), that the N400 effect elicited by coherent but prediction-inconsistent nouns critically hinges on wider discourse.

The ERPs elicited by prediction-consistent and -inconsistent nouns did in Experiment 2 diverge in a post-N400 latency range, from about 500 ms onward. We can offer only a very tentative explanation for this residual difference. As indicated for Cz in Figures 3 and 5, the ERP difference emerges in the latency range of estimated critical sentence offsets (as calculated relative to critical noun onset). On closer analysis, however, the two sets of nouns differed in how close they were to subsequent sentence offset, with the average discourse-predictable noun beginning 799 ms before sentence offset, and the average prediction-inconsistent noun beginning 866 ms before sentence offset. Because sentence offsets are usually associated with large ERP deflections, the late residual effect in Figure 5 is thus perhaps associated with an asymmetry in sentence offset timing. Alternatively, it might be associated with the fact that discourse-predictable nouns are on average somewhat shorter than prediction-inconsistent ones (5.68 vs. 6.72 phonemes), which implies that for nouns in nonfinal position, the next word and the associated ERP is shifted to the right. Whatever the exact cause of the late ERP differentiation observed in Figure 5, however, the most relevant observation remains as before: a sizable N400 effect elicited by prediction-inconsistent nouns in Experiment 1 but no such effect when the discourse context is removed in Experiment 2.

In Figure 6, we summarize the main ERP findings from Experiments 1 and 2 for a single electrode (RT). Panel A shows the ERPs time-locked to the estimated acoustic onset of the critical adjective in their discourse context for adjectives whose inflection is consistent or inconsistent with the discourse-predictable noun.² The waveforms in Panel A, although time-locked to adjective onset, clearly reveal around 1100–1600 ms the large N400 effect elicited by the later nouns (redisplayed with the appropriate time-locking to noun onset in Panel C). However, around 300–600 ms

² Relative to the corresponding Figure 1, the time scale of Panel A has been stretched to match that of the panels below it; note that the signal looks less “noisy” than it does in Figure 1, in which over 2 s of unfiltered signal is displayed in relatively time-compressed form.

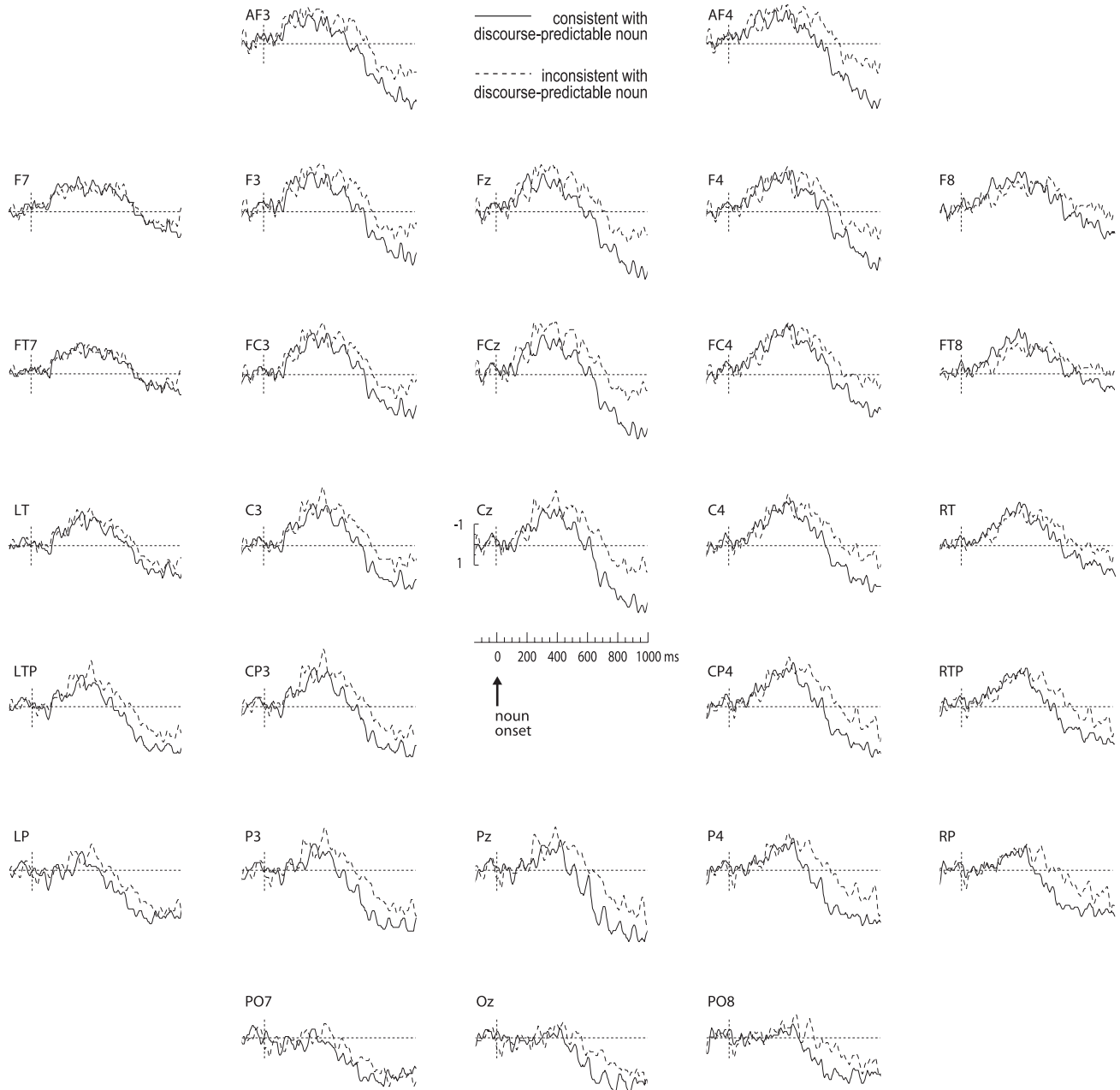


Figure 5. Nouns without prior discourse (Experiment 2). Grand average event-related brain potential waveforms time-locked to acoustic onset of formerly discourse-predictable nouns (solid) or coherent but formerly less predictable nouns (dotted) in their local carrier sentence.

from adjective onset, the waveforms in Panel A also reveal traces of the presently crucial adjective-elicited effect, the effect that comes out more clearly when time-locking to the estimated acoustic onset of the inflectional suffix in Panel B. As discussed before, we take the fact that this adjective effect “sharpens up” at estimated inflection onset as additional support for our account. Finally, Panels D and E display the results at estimated inflection and noun onset after removing the wider discourse in Experiment 2. They show that both the inflection-elicited early positive ERP

deflection and the noun-elicited N400 effect obtained in Experiment 1 critically hinged on the presence of that discourse.

Experiment 3

Although the details of the inflection-elicited EEG effect heavily depend on the fact that spoken language was used, nothing in our preferred explanation for this effect hinges on spoken language comprehension, on the use of EEG, or on the nature of

Table 5
Analyses of Variance (ANOVAs) on Mean Event-Related Brain Potential Amplitude in the 300–500 Milliseconds After Noun Onset in Nonpredictive Isolated Sentences (Experiment 2)

Source	<i>df</i>	<i>F</i>	<i>MSE</i>	<i>p</i>
Midline ANOVA (5 electrodes)				
PC	1, 23	1.17	14.00	.292
PC × EI	4, 92	0.40	1.96	.663
Quadrant ANOVA (2 × 2 × 5 electrodes)				
PC	1, 23	0.40	28.61	.531
PC × An	1, 23	0.29	8.67	.593
PC × He	1, 23	0.78	2.93	.387
PC × An × He	1, 23	0.08	0.60	.776
PC × An × He × EI	4, 92	1.05	0.18	.368
Simple main effects of prediction consistency for each electrode quadrant				
LA	1, 23	0.46	6.75	.503
RA	1, 23	0.00	8.03	.990
LP	1, 23	0.71	13.90	.408
RP	1, 23	0.28	12.11	.599

Note. For the midline and quadrant ANOVAs, only effects involving prediction consistency are reported: PC = prediction consistency (consistent and inconsistent); EI = electrode; An = anteriority (anterior and posterior); He = hemisphere (left and right). Also shown is the simple main effect of prediction consistency and the associated inconsistent-consistent amplitude difference in μV for each electrode quadrant: LA = left anterior; RA = right anterior; LP = left posterior; RP = right posterior. * $p < .05$.

the specific ERP effect at hand. If people anticipate specific upcoming nouns, and if this prediction is subsequently disconfirmed by a prenominal adjective whose inflection does not agree with the anticipated noun's gender, the processing costs of this unexpected turn of events might also show up in reading times. In Experiment 3, therefore, we presented a subset of our critical stories in a self-paced reading task.

Method

Subjects. For Experiment 3, we recruited 24 native speakers of Dutch (21 women and 3 men, mean age 21 years and range 18–33 years) from the student subject pool of the University of Amsterdam, Amsterdam, The Netherlands. None had participated in Experiments 1 or 2.

Materials. We selected 40 ministories from Experiment 1, of which 20 had a highly predictable neuter-gender noun and 20 a highly predictable common-gender noun. To accommodate potential spillover of effects in the self-paced reading task beyond the critical adjective, we modified the target sentences such that three words separated the critical first adjective from the later noun in all items (as in . . . *was situated behind a big but rather unobtrusive painting*). Across items, the average cloze probability of the discourse-predictable noun after the indefinite article was 89% ($SD = 5\%$) when the critical sentence was embedded in a wider discourse and 2% ($SD = 3\%$) when this sentence was presented in isolation. The critical items are available from www.josvanberkum.nl

In the first trial list, we presented half of the 40 critical stories (10 with a predictable common-gender noun and 10 with a neuter-gender one) in prediction-consistent form and the remainder in prediction-inconsistent form after matching the sets involved on mean cloze value of the discourse-

predictable noun in context and in isolation. The 40 critical stories were pseudorandomly mixed with 56 filler stories such that no more than 4 critical stories and no more than 2 critical stories in either consistency condition were immediately consecutive. The filler stories, of which 40 addressed a different issue (see Koornneef & Van Berkum, 2005), had an uncontrolled level of constraint. Each of three trial blocks began with two fillers, and they were preceded by a 10-story practice sequence. We derived the second list from the first one by rotating the condition of the critical items. We derived two more lists by reversing the order of the critical trials. Each list defined the session for 6 subjects, each of whom never saw an item in more than one condition.

Procedure and data analysis. We presented the stories in a standard noncumulative moving-window self-paced reading paradigm, using a non-proportional Courier 14-point font. Subjects read through each story word by word, with each button press disclosing the next word while replacing all other letters in the story with hyphens. As they pressed their way through a story, subjects could see its overall sentential and formatting layout (including punctuation) as well as the position of the currently visible word therein. To prevent edge effects in reading times, the critical region leading up to the noun was always separated from the left and right paragraph edges by at least one word. Subjects were asked to process each story for comprehension and to adapt their speed to this. Simple yes–no comprehension questions were asked after a pseudorandomly determined 50% of the stories. Comprehension questions that might focus the subject's attention on the research issue were avoided. A reading session consisted of four trial blocks separated by a short break, and took approximately 40 min on average.

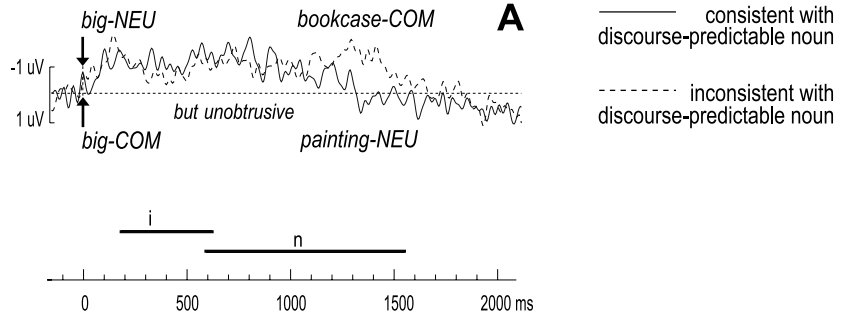
We analyzed word reading times in a region ranging from four words before the critical adjective up to and including the noun, referred to as $cw-4$ $cw-3$ $cw-2$ $cw-1$ adj $cw+1$ $cw+2$ $cw+3$ $noun$, respectively. For each of these 9 word positions, we computed mean reading time per subject and per item for each of two conditions (consistent and inconsistent with predictable noun) after eliminating all reading times that deviated more than 2 SD from both the mean reading time for the subject in that condition and the item in that condition (1.9% of the data, evenly distributed across the 9×2 cells of the design). We examined the prediction consistency effect in a by-subjects and a by-items ANOVA at each of the positions at which such an effect might show up (adj $cw+1$ $cw+2$ $cw+3$ $noun$) as well as at each of the positions at which it should not show up ($cw-4$ $cw-3$ $cw-2$ $cw-1$).

Results

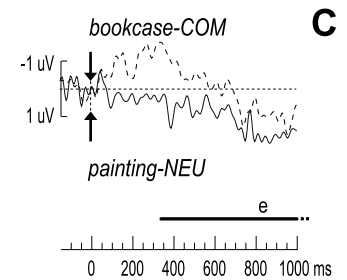
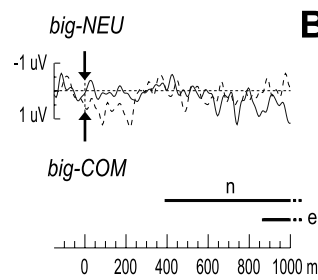
Across subjects, an average 94% ($SD = 6.2\%$) of the comprehension questions were answered correctly, with no subject falling below 75%. Table 6 displays mean reading times and the associated F statistics at nine word positions in sentences that were (at specific positions) consistent or inconsistent with the discourse-predictable noun, averaged across all 40 critical items in each condition and then averaged across the 24 subjects. As expected, there was no effect of prediction consistency at the four words leading up to the critical inflected adjective, where conditions do not yet differ, and a very large prediction consistency effect at the noun.

Unexpectedly, no clear inconsistency effect emerged at the inflected adjective or at the subsequent three words. However, there was an 18-ms trend toward delay at $cw+3$, the third word after the critical adjective. When we examined the set of actual words involved at this position, all of them turned out to be adjectives (as in . . . *was situated behind a big but rather unobtrusive painting*), and 37 of them carried a gender-marking inflectional suffix (as in the Dutch item in Example 3). In contrast, three adjectives at $cw+3$ did not carry a gender-marking inflectional

The burglar had no trouble whatsoever
to locate the secret family safe.
Of course, it was situated behind a...



The burglar had no trouble whatsoever
to locate the secret family safe.
Of course, it was situated behind a...



[no predictive discourse context]
Of course, it was situated behind a...

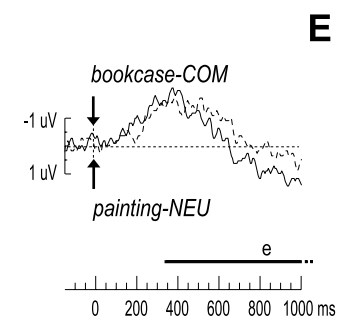
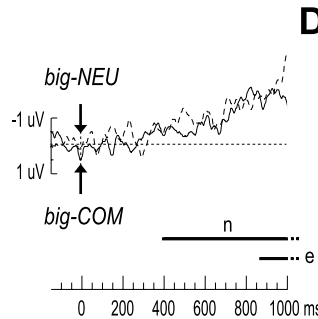


Figure 6. A summary of the event-related brain potential effects from Experiments 1 and 2. Panels A–E correspond to Figures 1–5, respectively, but zoom in on the data for single electrode, RT (see text for explanation).

suffix. These three adjectives, *plastic*, *geschreven*, and *gebroken* [plastic, written, and broken] are among a set of Dutch adjectives that never inflect for gender. To see whether they might have hidden a potentially reliable inflection inconsistency effect at the second adjective, we excluded the data for those three items and recomputed reading times. The results are shown in Table 7.

As before, there was no effect of prediction consistency before the critical first adjective or at the first adjective itself and the two words that immediately followed. However, readers did reliably slow down 21 ms at the inconsistently inflected second adjective (relative to the consistently inflected counterpart). And, as in the analysis shown in Table 6, they again slowed down considerably at the immediately consecutive inconsistent noun.

Discussion

In self-paced reading times, the processing consequences of a prediction-inconsistent adjective inflection did not emerge where we had seen them emerge in the spoken-language ERP experiment, that is, right at the first inconsistently inflected adjective. However, what is critical to our claim is that, although they emerged somewhat later, these processing consequences did emerge before the noun was seen. The only reasonable account for this finding is identical to that for the findings of Experiment 1, namely that (a) people can use discourse-level information to anticipate specific upcoming words as a sentence unfolds and (b) the syntactic gender properties of strongly anticipated nouns can immediately begin to

Table 6
Reading Time (in Milliseconds) Results Across All 40 Items in Experiment 3

Results	Word								
	cw-4 ...was	cw-3 situated	cw-2 behind	cw-1 a	Adj big-INFL	cw+1 but	cw+2 rather	cw+3 unobtrusive	Noun painting/bookcase
Reading times									
Consistent	408	365	360	327	342	350	373	407	498
Inconsistent	402	364	363	336	351	352	374	425	598
Effect size	-6	-1	3	9	9	2	1	18	100
F test									
$F_1(1, 23)$	0.46	0.03	0.24	3.05	2.46	0.07	0.02	3.36	15.32
$F_2(1, 38)$	0.28	0.07	0.55	1.77	1.31	0.06	0.04	4.10	23.94
MSE_1	1741	766	1033	515	779	1235	1466	2374	15854
MSE_2	2341	812	512	656	1076	927	1454	1672	8780
p_1	.504	.876	.632	.094	.131	.801	.879	.080	.001
p_2	.601	.790	.464	.191	.261	.812	.850	.050	.000

Note. cw-4-cw-1 = fourth to first word before the critical adjective; Adj = critical (first) adjective; cw + 1-cw + 3 = first to third word after the critical adjective (with cw + 3 being the second adjective); INFL = gender inflection.

interact with locally supplied syntactic constraints, in this case to disconfirm the specific anticipation.

As in Experiment 1, structured postsession interviews revealed that our subjects had not noticed the critical manipulation or associated stimulus characteristics. In part, this may again be due to the relative salience of certain aspects of the filler items (which contained relatively unexpected anaphoric references; cf. Koornneef & Van Berkum, 2005). However, as in Experiment 1, it also again suggests that the generation and subtle disconfirmation of strong lexical predictions does not in itself attract attention. And, specific to Experiment 3, it is therefore also unlikely that the present inconsistency effect emerged only several words after the first critical adjective because the subjects in this study somehow strategically exploited the fixed three-word distance between this critical adjective and the relevant noun.

As for why the reading time effect lags behind the ERP effect of Experiment 1, we can only speculate. One might argue, first of all, that the reading delay observed three words downstream from the first critical adjective is not elicited by the second adjective, but by

spillover from the first adjective. Spillover is frequently seen in self-paced reading (Mitchell, 2004) and is often attributed to the fact that subjects get into a relatively fixed button-press rhythm. In the present case, this would amount to assuming that subjects need three more button presses (~1,000 ms or more) before their surprise at the first adjective expresses itself in their response. Given the immediate large delay at prediction-inconsistent nouns, one would also have to argue—perhaps not unreasonably—that only very weak processing consequences suffer such long spillover. However, note that under a spillover account, there would be no reason, other than a pure coincidence, why the effect shows up only for items in which the second adjective also inflects for gender.

If we instead accept that the reading time effect is directly reflecting processes associated with the second adjective, this deeper delay also calls for an explanation. One possibility is that, perhaps because of the somewhat relaxed real-time constraints in self-paced reading as compared with speaker-paced listening, our readers engaged in discourse-based lexical prediction (and/or its

Table 7
Reading Time (in Milliseconds) Results Across 37 Items With Inflected 2nd Adjective at cw + 3 in Experiment 3

Results	Word								
	cw-4 ...was	cw-3 situated	cw-2 behind	cw-1 a	Adj big-INFL	cw+1 but	cw+2 rather	cw+3 (adj2) unobtrusive-INFL	Noun painting/bookcase
Reading times									
Consistent	403	364	359	327	344	349	368	405	487
Inconsistent	397	362	361	336	349	353	370	426	591
Effect size	-6	-2	2	9	5	4	2	21	104
F test									
$F_1(1, 23)$	0.66	0.31	0.13	2.46	0.81	0.26	0.08	4.50	19.08
$F_2(1, 35)$	0.39	0.35	0.21	2.04	0.42	0.20	0.12	5.84	21.69
MSE_1	1431	1129	1128	828	843	1324	1230	2404	13608
MSE_2	2639	929	549	652	1089	1031	1393	1405	9211
p_1	.424	.721	.727	.130	.379	.616	.775	.045	.000
p_2	.537	.559	.651	.162	.523	.666	.727	.021	.000

Note. cw-4-cw-1 = fourth to first word before the critical adjective; Adj = critical (first) adjective; cw + 1-cw + 3 = first to third word after the critical adjective (with cw + 3 being the second adjective); INFL = gender inflection.

syntactic verification) to a somewhat lesser degree than our listeners did. Alternatively, if the processing consequences of a disconfirmed lexical prediction happen to show up more clearly in ERPs than in reading times, obtaining a visible effect in the latter might require a stronger (e.g., double) disconfirmation. We cannot as yet discriminate between these various explanations. Whatever the exact cause of this difference in timing, though, both readers and listeners display evidence for discourse-based lexical anticipation.

General Discussion

In two ERP experiments and one self-paced reading study, we examined whether listeners and readers can use their knowledge of the wider discourse rapidly enough to anticipate specific upcoming words on the fly, as a sentence is unfolding. In the main ERP experiment (Experiment 1), subjects listened to Dutch stories that supported the prediction of a specific noun (e.g., *The burglar had no trouble locating the secret family safe. Of course, it was situated behind a . . .*). To probe whether listeners were indeed anticipating this noun (e.g., *painting*) by the time they had heard the indefinite article, critical stories were continued with a gender-marked adjective whose inflectional suffix did not agree with the noun's syntactic gender. Relative to consistently inflected adjectives, these prediction-inconsistent adjectives elicited a small but reliable positive deflection in the ERP waveforms, emerging right at the inflection. This ERP effect disappeared when subjects heard the same sentences without the prediction-supporting wider discourse (Experiment 2). Furthermore, when again presented in discourse in a self-paced reading study (Experiment 3), prediction-inconsistent adjectives also caused readers to slow down before the noun was shown.

The processing consequences reflected in ERPs and reading times in Experiments 1 and 3 were elicited by adjectives whose inflection did not agree with the lexically stored syntactic gender feature of a discourse-predictable noun. However, at the time at which these consequences were observed, the head noun had not yet been presented, and both of the adjective's inflectional variants were thus still fully grammatical. The only systematic difference between the inflected adjectives in our two critical conditions was whether their inflectional suffix agreed with the gender of the noun that was predictable at this point in the discourse. Therefore, the effects we observed in ERPs and reading times only make sense if, as we expected, our subjects had by this time indeed anticipated the discourse-predictable noun. In addition, because noun anticipation betrayed itself via an adjectival syntactic gender inflection, we can infer that syntactic features of an anticipated ghost noun are somehow involved in a syntactic analysis. We discuss these two central implications below and then briefly turn to the noun-elicited N400 effect.

Before doing so, we need to address a general concern that one might have over the level of constraint involved in our experiments. With average cloze values of 86% and 89% across items in Experiments 1 and 3, respectively, might our critical stories perhaps be unnaturally predictive to such an extent that the anticipatory processes observed in these experiments would not generalize to everyday language comprehension? Concerns over the validity of handcrafted "textoids" (Graesser, Millis, & Zwaan, 1997) should not be discarded lightly. In this particular case, though, we believe that the concern is unwarranted. To maximize the power of

our experiments, we carefully designed the critical stories to be highly predictive. However, we did so without compromising their naturalness or betraying their purpose. To avoid drawing attention to predictability and its disconfirmation via adjectival gender, for instance, we decided to avoid overt gender agreement violations by presenting the discourse-predictable noun after consistent adjectives only and by using an alternative-gender noun after inconsistent adjectives instead. The fact that our subjects failed to notice the critical manipulation even when prompted to comment on odd, regular, or annoying aspects of the materials suggests that we succeeded.

Note, furthermore, that the level of predictability reflected in average cloze values of 86% and 89% holds for a single specific position in each story only. At all other points in the story, the level of constraint was not controlled. It was also not controlled in the many filler stories. The extent to which our materials, on average, approach the level of constraint in natural language is difficult to establish. However, we suspect that with relatively decontextualized ministories such as these, the average level of predictability may well fall somewhat below the average level of constraint in, say, real-life conversation. In all, we see little reason to be concerned over this aspect of our study, and we are confident that the findings can be generalized to normal language use outside the laboratory. Whether upcoming words are only predicted at high-cloze story positions is a different issue and one to which we return below.

Discourse-Based Lexical Anticipation

The profound generativity of language might lead one to infer that the prediction of specific upcoming words is a doomed affair (e.g., Jackendoff, 2002) and is as such very unlikely to feature as part of the language comprehension system. Our findings clearly suggest otherwise. The reported ERP and self-paced reading experiments demonstrated that in a sufficiently constraining natural discourse, listeners and readers do predict specific upcoming words. Informally, such anticipation could already be observed in natural conversational exchanges, in which interlocutors can and do quite easily take over and finish each other's sentence. The results of Experiments 1 and 3 confirm and extend this observation. Most striking, the inflection-elicited ERP effect obtained in Experiment 1 reveals that listeners can anticipate specific words rapidly enough to affect the comprehension of fluently unfolding speech in midsentence, before the anticipated word comes along.

Our evidence for discourse-based lexical prediction converges with recent ERP evidence for sentence-based lexical prediction obtained by Wicha et al. (2004) with Spanish readers. Wicha et al. manipulated the gender of a prenominal determiner such that it did or did not agree with the gender of a sentence-predictable noun. Determiners that did not agree elicited a significant and slightly left-lateralized positivity in the ERPs around 500–700 ms after determiner onset. Because it was elicited by a formally correct prenominal gender marker (as in our experiments), this differential ERP effect strongly suggests that readers can use sentential context to predict specific upcoming words. Furthermore, the fact that both a single sentence and a somewhat larger discourse can induce such predictions suggests that this difference does not really matter, and that, as observed before for incremental interpretation (Van Berkum, Hagoort, & Brown, 1999; Van Berkum, Zwitserlood, et al.,

2003), the relevant interpretive context is simply the widest interpretive domain available.³

Our Experiment 1 differs from the Wicha et al. (2004) ERP experiment in language (Dutch vs. Spanish), input modality (spoken vs. written), input pacing (fully connected and naturally variable timing vs. fixed 500 ms per word presentation), source of the constraint (always involving prior discourse in our case vs. primarily sentential in the Wicha et al. study), gender probe type (adjectival suffix vs. various gender-marked determiners), and in whether the critical materials included overt gender agreement violations (they did not in our materials, but they did in half of the Wicha et al. materials). It remains to be seen which of these factors can explain why we obtained a slightly different effect. Note, however, that whereas the ERPs time-locked to inflection onset (see Figure 2) display a slightly right-lateralized reliable positivity between 50 and 250 ms, the corresponding trend observed in the ERPs time-locked to adjective onset (see Figure 1) bears some resemblance to the late positivity obtained by Wicha et al. (their Figure 6). Part of the reason for why our inflection-locked effect differs from the Wicha et al. finding might therefore be that, with spoken language, we were able to somewhat more precisely time-lock the unfolding EEG signals to the functionally critical stimulus, a disconfirming gender cue. We are currently examining this issue in our laboratory.⁴

Our ERP and reading time results are relevant to several domains of inquiry. Research on predictive inferences during text comprehension (e.g., Fincher-Kiefer, 1993; Graesser et al., 1994; Klin et al., 1999; McKoon & Ratcliff, 1992; Weingartner et al., 2003) has shown that people can use their knowledge of the situation described by the discourse to anticipate likely developments in that situation. Because we used a word's memorized lexicosyntactic gender to probe for discourse-based lexical prediction before the word itself comes along, our findings unequivocally reveal that people can also rapidly use their knowledge of the wider discourse to anticipate specific upcoming words. That is, with sufficient constraints, language users not only predict what might happen next in the world that is captured in their situation model but also what might happen next in the linguistic exchange at hand.

Our adjective-related findings also go beyond the classic predictability effects in written word recognition, such as that context-predictable words are read more quickly (e.g., Ehrlich & Rayner, 1981), responded to more quickly in naming and lexical decision tasks (e.g., Hess et al., 1995), and elicit smaller N400 effects (e.g., Kutas & Hillyard, 1984). The reading time advantage observed for discourse-predictable nouns in Experiment 3, and the reduced N400 elicited by those nouns in Experiment 1, replicate these well-attested phenomena. However, whereas these noun-elicited findings might reflect the consequences of discourse-based lexical anticipation, they may instead also emerge once the word at hand has been (at least partially) processed, because of an easier integration of the associated concept into the wider interpretive context (cf. Hess et al., 1995). Because our critical adjectives probe for anticipation before the noun, and do so via its lexically memorized and otherwise unpredictable gender feature, only our inflection-elicited processing effects provide direct evidence for lexical anticipation.

As such, our findings furthermore reveal that in contextually constraining context, the language comprehension system can go

beyond predicting the semantic features of upcoming words (Schwanenflugel & LaCount, 1988; Schwanenflugel & Shoben, 1985; Schwanenflugel & White, 1991), an idea that has been used to explain the constraint-dependent N400 effects discussed in the introduction to this article (cf. Federmeier & Kutas, 1999a, 1999b; Van Petten et al., 1999). What our results suggest is that, when faced with a sentence such as *John kept his gym clothes in a . . .*, native speakers of English do not just generate a set of semantic features that include, say, *small, rectangular, associated with gyms, holds clothes*, and *shutable* (Schwanenflugel & Shoben, 1985). Instead, or probably moreover, they actually predict the word *locker*.

The results of Experiments 1 and 3 can also be taken to bear on the principle of bottom-up priority in spoken and written word recognition. As mentioned before, models of word recognition that embody this principle (e.g., Forster, 1979, 1989; Marslen-Wilson, 1987, 1989; Norris, 1994) take a clear stance against prediction as being relevant to word recognition. However, the few studies that actually looked for context-based word preactivation and found no evidence for it (e.g., Zwitserlood, 1989) used only moderately constraining context, with cloze values around 20%–30%. We tested for lexical prediction with stories that were much more constraining (cloze values above 75%), and found clear evidence for such prediction. Note that our adjective-related effects, although revealing the processing consequences of anticipating an upcoming word at some level of the comprehension system, do not straightforwardly tap the processes involved in recognizing the noun. Furthermore, we do not wish to claim that the lexical preactivation uncovered in our studies is accompanied by hallucinations about an actual noun being presented in the input. Nevertheless, it seems reasonable to assume that the word recognition system will benefit by the time the predicted noun actually does come along (perhaps reflected, to an unknown extent, by the faster reading of predictable words, cf. Experiment 3). Moreover, in the unlikely event that word recognition itself would not gain from a correct lexical prediction (and suffer from an incorrect one), other aspects of the comprehension system—parsing and interpretation—are bound to do so.

What mechanisms are involved in generating these lexical predictions? The results of Experiment 2 showed that the lexical anticipation observed critically hinged on the wider discourse. However, this cannot be all there is to it. After all, whereas the

³ In the Wicha et al. (2004) experiment, the critical sentences were also embedded in a two-sentence ministry. However, the relation between critical and noncritical sentences was not controlled such that the degree of constraint critically hinged on prior discourse. Furthermore, approximately half of the critical sentences were in story-initial position. We therefore take their study to primarily inform us about the impact of sentential constraint.

⁴ In two earlier ERP experiments that mixed constraining sentence fragments with pictures instead of nouns (Wicha, Bates, Moreno, & Kutas, 2003; Wicha, Moreno, & Kutas, 2003), prediction-inconsistent Spanish articles elicited an N400-like negativity. It is not entirely clear why the ERP effects in these experiments are different from those involving natural language only. However, the mixing of language and pictures (as well as the consequently more transparent experimental manipulation) may well have affected how subjects in these studies processed the linguistic input (see Wicha et al., 2004, p. 1285, for a comparable suggestion).

prior discourse context may suggest the relevance of particular concepts (theft, jewelry, painting, etc.), it only makes sense to predict specific upcoming words—let alone to evaluate their syntactic features—in the context of an unfolding sentence. Moreover, which of the many discourse-relevant concepts will potentially soon be verbalized by the speaker will normally also hinge on the structure and content of this unfolding sentence. In *Of course, it was situated behind a*, for instance, it is the syntactic structure, in particular the indefinite article, that unequivocally signals that a head noun is bound to show up soon. Furthermore, it is the semantics of words like *situated*, *behind*, and *of course* that, in the wider context at hand, ultimately suggest that this noun is probably going to be *painting*.

In line with earlier findings on nonpredictive incremental interpretation (Van Berkum, Hagoort, & Brown, 1999; Van Berkum, Zwitserlood, et al., 2003), we suggest that the interpretive context that allowed our subjects to anticipate the word *painting* is a single unified model of the discourse and the situation described (Kintsch, 1998; Zwaan, 1999) that includes the semantic contribution made so far by the unfolding sentence. Furthermore, we propose that it is the rapid word-by-word combination of this continuously updated set of interpretive constraints with local syntactic (and, in speech, phonological) constraints that in the end supports the prediction of specific upcoming words. Moreover, because such prediction is about communication, we suspect that knowledge of the speaker and the common ground between speaker and listener will also play a role. After all, it makes little sense to predict that a 5-year old child will continue his or her currently unfolding utterance with *globalization*, even if the topic of discourse involves things that happen all over the world.

These ideas fit well with recent evidence indicating that the language comprehension system can use verb information to predict specific arguments in a variety of wider contexts (Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003; Kamide, Scheepers, & Altmann, 2003; Nieuwland & Van Berkum, 2004). In some of these studies, verb-supported predictions recruit information supplied by a prior discourse. In others, they exploit information supplied by a nontextual scene. In line with these findings, and with the perspective outlined in Van Berkum, Zwitserlood, et al. (2003), we suggest that predictions about how an utterance will unfold can draw upon information from any relevant interpretive domain—the prior discourse, a scene, a much earlier conversation, general world knowledge, cospeech gestures, inferred characteristics of the speaker, and so forth—as long as this information is made relevant or recruited by locally unfolding constraints (*Of course, it was hidden behind a . . .*). That is, any context made relevant by the currently unfolding sentence will do.

One might argue that, instead of being based on a deep message-level representation of the context up to and including the currently unfolding sentence, the lexical anticipation observed in Experiments 1 and 3 perhaps involves some form of convergent priming from multiple words in the preceding text (*burglar, safe, hidden*). This would be in line with the combination priming account proposed for text-based predictability effects (Duffy, Henderson, & Morris, 1989; see also the comparable concept of lexicosemantic fit proposed by Hoeks, Stowe, & Doedens, 2004). Although we were able to avoid strong lexical associates in many of our context sentences, we cannot exclude the possibility that part or all of the adjective-elicited effects hinge on some subtle form of combina-

tion priming. In our perspective, this would by no means make the phenomenon less interesting—only an exaggerated focus on classic modularity issues would do so. However, we currently favor the message-level account. The main reason is that in Dutch, the detection of an adjective-noun gender agreement violation requires some nonlocal syntactic parsing. In particular, Dutch prenominal adjectives mark their head noun's gender in indefinite singular noun phrases only and not in definite singular NPs or plural NPs. So, for instance, whereas *painting_{neu}* is associated with the zero inflection \emptyset in a singular indefinite NP, it is associated with the *-e* inflection in singular definite NPs. It seems unlikely that a parsing system capable of handling this type of nonlocal inflectional complexity would be grafted right on top of a simple convergent intralexical priming mechanism. In our laboratory, we are currently investigating whether this “argument from design” is pointing in the right direction (see Townsend & Bever, 2001, for an interesting alternative).

Turning to another aspect of the mechanism, it is as yet unclear when and how the informational input for a discourse-based lexical prediction is actually converted into a concrete prediction. One possibility is that whenever the syntactic and interpretive context is sufficiently constraining, the system makes a discrete prediction of a single word that is singled out because its probability has exceeded some absolute or relative threshold. In the spirit of connectionist models of language comprehension (cf. Elman, 1990, 1995; Seidenberg & MacDonald, 1999; Tabor, Juliano, & Tanenhaus, 1997; Tabor & Tanenhaus, 1999), another possibility is that the system continually makes *graded predictions*, as such defining a probability landscape over the entire lexicon. Our data do not allow us to decide between these two accounts. However, because a discrete two-step mechanism would need to keep track of graded probabilities too, the graded prediction account should perhaps be preferred on grounds of parsimony. The latter can also elegantly explain why, when contextual constraints lead to only moderate predictability (e.g., Zwitserlood, 1989), some word-initial input is needed to bring these constraints to bear.

A potentially viable third way to conceive of discourse-based lexical prediction is in terms of covert language production. It has recently been suggested (Garrett, 2000; Jackendoff, 2002; Kempen, 2000; Pickering & Garrod, 2004; Townsend & Bever, 2001) that language perceivers can recruit parts of their language production system to very rapidly resolve ambiguity (*If I were the speaker, which of the competing alternative readings of the input that I'm hearing right now would I have produced myself?*). This opens up the interesting possibility that they might also be able to recruit parts of the language production system to make discourse-based lexical predictions by essentially asking themselves *If I were the speaker, what would I say next?* Note that at any point in their unfolding utterance, speakers usually come up with one word only. In contrast with the discrete and graded prediction mechanisms discussed above, therefore, discourse-based lexical prediction via the regular mechanism for lemma access in speech production (Levelt, 1989; Levelt, Roelofs, & Meyer, 1999) would be naturally constrained to generating just a single specific word and to doing so in sufficiently constraining contexts only, as such relieving the syntactic parser—as well as other levels of the comprehension system that take projected structure into account—from having to deal with a multitude of projected analyses. Of course, for such anticipation to work, it should not require the level of attention

needed for normal preverbal message planning in speech production. However, with much of the preverbal message already in place as a result of incremental comprehension, the system may be off on a good start.

Continuous Syntax-Based Evaluation

Beyond demonstrating rapid discourse-based lexical anticipation, our findings have a very interesting second implication. We know from other EEG research that Dutch listeners and readers rapidly detect gender agreement violations, both when the violation is overt and impossible to repair (as in *a big_{com} painting_{neu}*; e.g., Bastiaansen, Van Berkum, & Hagoort, 2002; Hagoort, 2003b; Hagoort & Brown, 1999; Van Berkum, Zwitserlood, Brown, & Hagoort, 2000) as well as when the violation hinges on the provisional commitment to a particular syntactic parse (Van Berkum, Brown, & Hagoort, 1999a, 1999b). The current findings reveal that people can, in at least some natural circumstances, also detect agreement violations involving an anticipated head noun. As outlined before, the mapping between a Dutch head noun's gender and an adjectival inflection is not straightforward and requires some nontrivial syntactic analysis process. Our findings suggest that this process is partly anticipatory, in that it can relate its incremental parse of the unfolding sentence to the syntactic features of anticipated ghost nouns.

The abovementioned EEG research on Dutch gender agreement violations has shown that these violations invariably elicit a so-called P600/syntactic positive shift (SPS) effect, an ERP effect that is more generally associated with syntactic parsing problems (Hagoort, Brown & Groothusen, 1993; Osterhout & Holcomb, 1992; see Brown, Hagoort, & Kutas, 2000, or Hagoort, Brown & Osterhout, 1999, for review). Although the classic P600/SPS effect is known to vary in scalp distribution as well as onset (see Hagoort, 2003a, for an example of the latter), the positivity we obtain at prediction-inconsistent adjective inflections falls outside the range typically observed, particularly in terms of its onset. This could be taken to indicate that the processing consequences of overt and anticipated syntactic gender violations are functionally distinct. Note, however, that the classic benchmark effect is always computed relative to word onset. Without knowing what the relevant spoken-language effects would look like when time-locked to the critical information within the spoken word, all we can do is compare the extant P600/SPS effects with the waveforms locked to adjective onset in Figure 1. Taking the somewhat high time compression of this figure into account (see Figure 6A for a more conventional close-up at RT), the differential trend observed in the ~300–1,000-ms latency range is not unlike other reported P600/SPS effects. We must thus accept that no firm conclusion can as yet be drawn here.

Why would the comprehension system engage in such prediction-sensitive parsing and complicate matters by relating its incremental syntactic analysis to not just overtly presented words, but to anticipated ghost words as well? In a system in which the incrementally constructed interpretive, syntactic, and phonological representations of linguistic input are tightly linked (Jackendoff, 1999, 2002; MacDonald, Pearlmutter, & Seidenberg, 1994; Tanenhaus & Trueswell, 1995), a prediction made at one level of representation can easily lead to an associated prediction at another level. In Jackendoff's (2002) framework, in which the lexical

representation of a word is defined as an idiosyncratic coupling of fragments of phonological, syntactic, and conceptual structure, word prediction would amount to the prediction of upcoming phonological, syntactic, and conceptual structure. Of course, the comprehension system may choose to selectively attend to particular aspects of this prediction only by, for example, focusing on predicted phonological form while consistently ignoring anticipated bits of syntax. However, the best way to keep a predictive linguistic system in check and prevent it from being bogged down in too many (or internally inconsistent) predictions is to have it continuously verify and adjust its predictions by using constraints at all a priori relevant layers of information.

This broader perspective actually suggests an alternative reading of the ERP and reading time effects elicited by prediction-inconsistent adjective inflections. Instead of the direct processing consequences of a gender agreement violation between anticipated nouns and subsequently mismatching adjectives, these effects may also reflect the processing consequences associated with adjusting a prediction. After all, although we designed our items to generate a strong prediction at the indefinite article, the language comprehension system of course does not know about this privileged point. All this system can know about is the specific information that any given word brings along with it, and the new predictions, if any, that can be made from there.

Discourse-Based N400 Effect

The ERP waveforms elicited by the critical later nouns in Experiments 1 and 3 corroborate and extend prior evidence on the sensitivity of the word-elicited N400 to discourse-level factors (e.g., Federmeier & Kutas, 1999a, 1999b; St. George et al., 1994; Van Berkum, Hagoort, & Brown, 1999; Van Berkum, Zwitserlood, et al., 2003). In line with these earlier findings, the data displayed in Figures 3 and 5 suggest that words are extremely rapidly related to a representation of what the wider discourse is about. Note that in contrast with previously reported discourse modulations of the N400, the discourse-dependent N400 effect obtained in Experiment 1 does not hinge on outright semantic anomaly. Instead, what we see here may well be the discourse-level equivalent of what has been observed before in single sentence research (e.g., Hagoort & Brown, 1994; Kutas & Hillyard, 1984): A coherent but nevertheless somewhat less preferred word like *bookcase* elicits a larger N400 than a coherent and highly preferred word like *painting*.

As outlined before, it is difficult to tell whether the reduced N400 amplitudes elicited by discourse-predictable words are due to actual lexical prediction or to facilitated postlexical integration.⁵ However, the effect shown in Figure 3 is somewhat ambiguous in another respect as well. The reason is that, to the extent that the comprehension system must recover from a strong prediction (*painting*) having been disconfirmed by a preceding adjective

⁵ Other evidence (Van Berkum, Zwitserlood, et al., 2003, Figure 3) does suggest that discourse-dependent N400 anomaly effects do not depend on whether the anomalous word disconfirms a strong lexical prediction. In particular, discourse-anomalous spoken words also elicited a large N400 effect relative to a very low-cloze (<5%) discourse-coherent control word. This speaks in favor of a postlexical integration account of N400 effects in language comprehension (in line with, e.g., Brown & Hagoort, 1993; Chwilla, Kolk, & Mulder, 2000).

inflection, the processing consequences of this unexpected turn of events perhaps also affect the response to the less-preferred noun (*bookcase*). As the preferred noun (*painting*) is always preceded by a consistent adjective, its processing will not suffer from such earlier disconfirmation. This confounding factor is perhaps responsible for the strikingly early onset of the N400 effect at hand. Hence, although certainly consistent with earlier findings (e.g., Van Berkum, Zwitserlood, et al., 2003), the discourse-dependent N400 effect reported here should be interpreted with care. We are currently investigating, without a confounding prenominal adjective, whether coherent nouns in stories like these elicit differential N400 effects as a function of how preferred they are (see Otten & Van Berkum, 2005, for initial results).

Conclusions

Anticipation plays a vital role in many aspects of our lives. The evidence from ERPs and reading times presented here suggest that language comprehension is no exception. We have shown, first of all, that listeners and readers anticipate upcoming words in discourse. They do not just do it when their conversational partner hesitates or when a written sentence terminates prematurely in a cloze test. Words can be anticipated on the fly, as fluent discourse unfolds. Second, because we detected the anticipation of specific nouns via broken syntactic gender agreement between this noun and a prenominal adjective, our findings suggest that the language comprehension system engages in prediction-sensitive parsing, relating the syntactic features of expected—but not yet presented—words to an incremental syntactic analysis of the sentence presented so far. These two findings are probably deeply related, for a system that continually adjusts its predictions in the face of new evidence will do better that one that does not. In all, our findings show that anticipation or prediction is a pervasive aspect of ordinary language comprehension, affecting several levels of the comprehension system involved.

To avoid being misconstrued: we do not conceive of understanding a sentence in terms of predicting what word or associated syntactic structure will come next. Comprehension must in the end work with what it has, not with what it believes will be. However, as laid out in the introduction, real language use is not a random affair solely constrained by uncoupled sets of grammatical rules. Different layers of the grammar are tightly linked to each other through the lexicon, as well as through correlated constraints (MacDonald et al., 1994; Tanenhaus & Trueswell, 1995) or interface rules (Jackendoff, 2002) that express mappings at a supra-lexical level. In addition, the interpretive layer is strongly correlated to the real world outside the language system. Although the comprehension system must in the end face what was actually said, this does not mean it cannot naturally exploit all this wisdom to anticipate a little. After all, predicting the trajectory of a frisbee does not preclude actually catching it. It helps.

References

- Altmann, G. T. M. (1997). *The ascent of Babel: An exploration of language, mind, and understanding*. Oxford, England: Oxford University Press.
- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.
- Bastiaansen, M. C. M., Van Berkum, J. J. A., & Hagoort, P. (2002). Syntactic processing modulates the theta rhythm of the human EEG. *NeuroImage*, *17*, 1479–1492.
- Brown, C. M., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience*, *5*, 34–44.
- Brown, C. M., Hagoort, P., & Kutas, M. (2000). Postlexical integration processes in language comprehension: Evidence from brain-imaging research. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 881–895). Cambridge, MA: MIT Press.
- Calvo, M. G. (2001). Working memory and inferences: Evidence from eye fixations during reading. *Memory*, *9*, 365–381.
- Calvo, M. G., Meseguer, E., & Carreiras, M. (2001). Inferences about predictable events: Eye movements during reading. *Psychological Research*, *65*, 158–169.
- Campion, N., & Rossi, J. P. (2001). Associative and causal constraints in the process of generating predictive inferences. *Discourse Processes*, *31*, 263–291.
- Chomsky, N. (1957). *Syntactic structures*. Den Haag, The Netherlands: Mouton.
- Chwilla, D. J., Kolk, H. H. J., & Mulder, G. (2000). Mediated priming in the lexical decision task: Evidence from event-related potentials and reaction time. *Journal of Memory and Language*, *42*, 314–341.
- Connine, C. M. (1987). Constraints on interactive processes in auditory word recognition: The role of sentence context. *Journal of Memory and Language*, *26*, 527–538.
- Connine, C. M. (1990). Effects of sentence context and lexical knowledge in speech processing. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing* (pp. 280–294). Cambridge, MA: MIT Press.
- Cutler, A., & Clifton, C. E. (1999). Comprehending spoken language: A blueprint of the listener. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 123–166). Oxford, England: Oxford University Press.
- Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, *16*, 507–534.
- Dahan, D., & Tanenhaus, M. K. (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.
- Duffy, S. A., Henderson, J. M., & Morris, R. K. (1989). Semantic facilitation of lexical access during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 791–801.
- Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, *20*, 641–655.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 197–211.
- Elman, J. L. (1995). Language as a dynamical system. In R. F. Port & T. van Gelder (Eds.), *Mind as motion: Explorations in the dynamics of cognition* (pp. 195–223). Cambridge, MA: MIT Press.
- Federmeier, K. D., & Kutas, M. (1999a). Right words and left words: Electrophysiological evidence for hemispheric differences in meaning processing. *Cognitive Brain Research*, *8*, 373–392.
- Federmeier, K. D., & Kutas, M. (1999b). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*, 469–495.
- Federmeier, K. D., McLennan, D. B., De Ochoa, E., & Kutas, M. K. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, *39*, 133–146.
- Fincher-Kiefer, R. (1993). The role of predictive inferences in situation model construction. *Discourse Processes*, *16*, 99–124.

- Fincher-Kiefer, R. (1995). Relative inhibition following the encoding of bridging and predictive inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 981–995.
- Fincher-Kiefer, R. (1996). Encoding differences between bridging and predictive inferences. *Discourse Processes*, 22, 225–246.
- Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 27–85). Hillsdale, NJ: Erlbaum.
- Forster, K. I. (1989). Basic issues in lexical processing. In W. D. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 75–107). Cambridge, MA: MIT Press.
- Foss, D.J. (1982). A discourse on semantic priming. *Cognitive Psychology*, 14, 590–607.
- Frazier, L. (1999). *On sentence interpretation*. Dordrecht, The Netherlands: Kluwer.
- Garrett, M. F. (2000). Remarks on the architecture of language processing systems. In Y. Grodzinsky, L. Shapiro, & D. Swinney (Eds.), *Language and the brain* (p. 31). San Diego, CA: Academic Press.
- Gaskell, M. G., & Marslen-Wilson, W. D. (2001). Lexical ambiguity resolution and spoken word recognition: Bridging the gap. *Journal of Memory and Language*, 44, 325–349.
- Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology*, 48, 163–189.
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101, 371–395.
- Grice, P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). New York: Seminar Press.
- Grosjean, F. (1980). Spoken word recognition and the gating paradigm. *Perception & Psychophysics*, 28, 267–283.
- Hagoort, P. (2003a). How the brain solves the binding problem for language: A neurocomputational model of syntactic processing. *NeuroImage*, 20, S18–S29.
- Hagoort, P. (2003b). The interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience*, 15, 883–889.
- Hagoort, P., & Brown, C. M. (1994). Brain responses to lexical-ambiguity resolution and parsing. In C. Clifton Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 45–80). Hillsdale, NJ: Erlbaum.
- Hagoort, P., & Brown, C. M. (1999). Gender electrified: ERP evidence on the syntactic nature of gender processing. *Journal of Psycholinguistic Research*, 28, 715–728.
- Hagoort, P., Brown, C. M., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, 8, 439–483.
- Hagoort, P., Brown, C. M., & Osterhout, L. (1999). The neurocognition of syntactic processing. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 273–316). Oxford, England: Oxford University Press.
- Harley, T. (2001). *The psychology of language: From data to theory*. Hove, England: Psychology Press.
- Hess, D. J., Foss, D. J., & Carroll, P. (1995). Effects of global and local context on lexical processing during language comprehension. *Journal of Experimental Psychology: General*, 124, 62–82.
- Hoeks, J. C., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: The interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, 19, 59–73.
- Jackendoff, R. (1999). The representational structures of the language faculty and their interactions. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 37–79). Oxford, England: Oxford University Press.
- Jackendoff, R. (2002). *Foundations of language*. New York: Oxford University Press.
- Jay, T. B. (2003). *The psychology of language*. Upper Saddle River, NJ: Prentice Hall.
- Jongman, A. (1998). Effects of vowel length and syllabic structure on segment duration in Dutch. *Journal of Phonetics*, 26, 207–222.
- Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49, 133–156.
- Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, 32, 37–55.
- Keefe, D. E., & McDaniel, M. A. (1993). The time course and durability of predictive inferences. *Journal of Memory and Language*, 32, 446–463.
- Kempen, G. A. M. (2000). Could grammatical encoding and grammatical decoding be subserved by the same processing module? *Behavioral and Brain Sciences*, 23, 38–39.
- Kemps, R., Ernestus, M., Schreuder, R., & Baayen, R. H. (in press). Prosodic cues for morphological complexity: The case of Dutch plural nouns. *Memory & Cognition*.
- Kemps, R., Wurm, L., Ernestus, M., Schreuder, R., & Baayen, R. H. (2005). Prosodic cues for morphological complexity: Comparatives and agent nouns in Dutch and English. *Language and Cognitive Processes*, 20, 43–74.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, England: Cambridge University Press.
- Kleiman, G. M. (1980). Sentence frame contexts and lexical decisions: Sentence-acceptability and word-relatedness effects. *Memory & Cognition*, 8, 336–344.
- Klin, C. M., Guzman, A. E., & Levine, W. H. (1999). Prevalence and persistence of predictive inferences. *Journal of Memory and Language*, 40, 593–604.
- Koornneef, A. W., & Van Berkum, J. J. A. (2005). *On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking*. Manuscript submitted for publication.
- Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 12, 463–470.
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307, 161–163.
- Levelt, W. J. M. (1989). *Speaking*. Cambridge, MA: MIT Press.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–38.
- Linderholm, T. (2002). Predictive inference generation as a function of working memory capacity and causal text constraints. *Discourse Processes*, 34, 259–280.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676–703.
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word recognition. *Cognition*, 25, 71–102.
- Marslen-Wilson, W. D. (1989). Access and integration: Projecting sound onto meaning. In W. D. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 3–24). Cambridge, MA: MIT Press.
- Marslen-Wilson, W. D., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, 8, 1–71.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86.

- McClelland, J. L., & O'Regan, J. K. (1981). Expectations increase the benefit derived from parafoveal visual information in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 634–644.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, 88, 375–407.
- McDonald, S. A., & Shillcock, R. C. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science*, 14, 648–652.
- McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, 99, 440–466.
- Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90, 227–234.
- Mitchell, D. C. (1994). Sentence parsing. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 375–409). New York: Academic Press.
- Mitchell, D. C. (2004). On-line methods in language processing: Introduction and historical review. In M. Carreiras & C. Clifton Jr. (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERPs and beyond* (pp. 15–32). New York: Psychology Press.
- Morris, R. K. (1994). Lexical and message-level sentence context effects on fixation times in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 92–103.
- Morris, R. K., & Folk, J. R. (1998). Focus as a contextual priming mechanism. *Memory & Cognition*, 26, 1313–1322.
- Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, 76, 165–178.
- Murray, J. D., & Burke, K. A. (2003). Activation and encoding of predictive inferences: The role of reading skill. *Discourse Processes*, 35, 81–102.
- Murray, J. D., Klin, C. M., & Myers, J. L. (1993). Forward inferences in narrative text. *Journal of Memory and Language*, 32, 464–473.
- Nieuwland, M., & Van Berkum, J. J. A. (2004, April). *Discourse context can completely overrule lexical-semantic violations: Evidence from the N400*. Paper presented at the annual meeting of the Cognitive Neuroscience Society, San Francisco.
- Nooteboom, S. G. (1972). *Production and perception of vowel duration: A study of durational properties of vowels in Dutch*. Unpublished doctoral dissertation, University of Utrecht, The Netherlands.
- Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52, 189–234.
- O'Regan, J. K. (1979). Eye guidance in reading: Evidence for the linguistic control hypothesis. *Perception & Psychophysics*, 25, 501–509.
- Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31, 785–806.
- Otten, M., & Van Berkum, J. J. A. (2005, April). *The influence of message-based predictability and lexical association on the discourse-dependent N400 effect*. Poster presented at the 2005 Annual Meeting of the Cognitive Neuroscience Society, New York.
- Perfetti, C. A. (1999). Comprehending written language: A blueprint of the reader. In P. Hagoort & C. M. Brown (Eds.), *Neurocognition of language processing* (pp. 167–208). Oxford, England: Oxford University Press.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27, 1–22.
- Pinker, S. (1994). *The language instinct*. New York: HarperCollins.
- Sachs, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language*, 50, 696–735.
- 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.
- Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, 110, 474–494.
- Samuel, A. G. (1990). Using perceptual-restoration effects to explore the architecture of perception. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing* (pp. 295–314). Cambridge, MA: MIT Press.
- Schmalhofer, F., McDaniel, M. A., & Keefe, D. (2002). A unified model for predictive and bridging inferences. *Discourse Processes*, 33, 105–132.
- Schwanenflugel, P. J., & LaCount, K. L. (1988). Semantic relatedness and the scope of facilitation for upcoming words in sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 344–354.
- Schwanenflugel, P. J., & Shoben, E. J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language*, 24, 232–252.
- Schwanenflugel, P. J., & White, C. R. (1991). The influence of paragraph information on the processing of upcoming words. *Reading Research Quarterly*, 26, 160–177.
- Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, 23, 569–588.
- St. George, M., Mannes, S., & Hoffman, J. E. (1994). Global semantic expectancy and language comprehension. *Journal of Cognitive Neuroscience*, 6, 70–83.
- Swinney, D. (1979). Lexical access during sentence comprehension: (Re) consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 18, 645–660.
- Tabor, W., Juliano, C., & Tanenhaus, M. K. (1997). Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes*, 12, 211–271.
- Tabor, W., & Tanenhaus, M. K. (1999). Dynamical models of sentence processing. *Cognitive Science*, 23, 491–515.
- Tanenhaus, M. K., Carlson, G. N., & Seidenberg, M. S. (1985). Do listeners compute linguistic representations? In D. R. Dowty, L. Karttunen, & A. M. Zwicky (Eds.), *Natural language processing: Psychological, computational, and theoretical perspectives*. New York: Cambridge University Press.
- Tanenhaus, M. K., & Trueswell, C. (1995). Sentence comprehension. In J. L. Miller & P. D. Eimas (Eds.), *Speech, language, and communication* (pp. 217–262). San Diego, CA: Academic Press.
- Townsend, D. J., & Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules*. Cambridge, MA: MIT Press.
- Traxler, M. J., & Foss, D. J. (2000). Effects of sentence constraint on priming in natural language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1266–1282.
- Traxler, M. J., Foss, D. J., Seely, R. E., Kaup, B., & Morris, R. K. (2000). Priming in sentence processing: Intralexical spreading activation, schemas, and situation models. *Journal of Psycholinguistic Research*, 29, 581–595.
- Van Berkum, J. J. A. (1996). *The psycholinguistics of grammatical gender: Studies in language comprehension and production*. Unpublished doctoral dissertation, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands. [chap. 2 is downloadable from www.josvanberkum.nl]
- Van Berkum, J. J. A. (2004). Sentence comprehension in a wider discourse: Can we use ERPs to keep track of things? In M. Carreiras & C. Clifton Jr. (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERPs and beyond* (pp. 229–270). New York: Psychology Press.
- Van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999a). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, 41, 147–182.
- Van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999b). When does

gender constrain parsing? Evidence from ERPs. *Journal of Psycholinguistic Research*, 28, 555–571.

Van Berkum, J. J. A., Brown, C. M., Hagoort, P., & Zwitserlood, P. (2003). Event-related brain potentials reflect discourse-referential ambiguity in spoken-language comprehension. *Psychophysiology*, 40, 235–248.

Van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, 11, 657–671.

Van Berkum, J. J. A., Zwitserlood, P., Bastiaansen, M. C. M., Brown, C. M., & Hagoort, P. (2004, April). *So who's "he" anyway? Differential ERP and ERSP effects of referential success, ambiguity and failure during spoken language comprehension*. Paper presented at the annual meeting of the Cognitive Neuroscience Society, San Francisco.

Van Berkum, J. J. A., Zwitserlood, P., Brown, C. M., & Hagoort, P. (2000, September). *The computation of gender and number agreement in parsing: An ERP-based comparison*. Paper presented at the 6th Conference on Architectures and Mechanisms of Language Processing, Leiden, The Netherlands.

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

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

Weingartner, K. M., Guzmán, A. E., Levine, W. H., & Klin, C. M. (2003). When throwing a vase has multiple consequences: Minimal encoding of predictive inferences. *Discourse Processes*, 36, 131–146.

Wennerstrom, A., & Siegel, A. F. (2003). Keeping the floor in multiparty conversations: Intonation, syntax, and pause. *Discourse Processes*, 36, 77–107.

Whitney, P. (1998). *The psychology of language*. Boston: Houghton Mifflin.

Whitney, P., Ritchie, B. G., & Crane, R. S. (1992). The effect of foregrounding on readers' use of predictive inferences. *Memory & Cognition*, 20, 424–432.

Wicha, N. Y., Bates, E. A., Moreno, E. M., & Kutas, M. (2003). Potato not pope: Human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters*, 346, 165–168.

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2003). Expecting gender: An event related brain potential study on the role of grammatical gender in comprehending a line drawing within a written sentence in Spanish. *Cortex*, 39, 483–508.

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, 16, 1272–1288.

Zwaan, R. A. (1999). Situation models: The mental leap into imagined worlds. *Current Directions in Psychological Science*, 8, 15–18.

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

Zwitserlood, P. (1998). Spoken words in sentence contexts. In A. D. Friederici (Ed.), *Language comprehension: A biological perspective*. Heidelberg/Berlin: Springer.

Zwitserlood, P. (2004). Sublexical and morphological information in speech processing. *Brain and Language*, 90, 368–377.

Received July 12, 2004

Revision received October 19, 2004

Accepted November 18, 2004 ■

ORDER FORM

Start my 2005 subscription to *Journal of Experimental Psychology: Learning, Memory, and Cognition!* ISSN: 0278-7393

___ \$140.00, APA MEMBER/AFFILIATE _____

___ \$287.00, INDIVIDUAL NONMEMBER _____

___ \$700.00, INSTITUTION _____

In DC add 5.75% / In MD add 5% sales tax _____

TOTAL AMOUNT ENCLOSED \$ _____

Subscription orders must be prepaid. (Subscriptions are on a calendar year basis only.) Allow 4-6 weeks for delivery of the first issue. Call for international subscription rates.



AMERICAN
PSYCHOLOGICAL
ASSOCIATION

SEND THIS ORDER FORM TO:
American Psychological Association
Subscriptions
750 First Street, NE
Washington, DC 20002-4242

Or call (800) 374-2721, fax (202) 336-5568.
TDD/TTY (202) 336-6123.
For subscription information, e-mail:
subscriptions@apa.org

Send me a FREE Sample Issue

Check enclosed (make payable to APA)

Charge my: VISA MasterCard American Express

Cardholder Name _____

Card No. _____ Exp. Date _____

Signature (Required for Charge)

BILLING ADDRESS:

Street _____

City _____ State _____ Zip _____

Daytime Phone _____

E-mail _____

SHIP TO:

Name _____

Address _____

City _____ State _____ Zip _____

APA Member # _____ *ALMA15*