The Neural Integration of Speaker and Message

Jos J. A. Van Berkum[1,2,3], Danielle van den Brink[3,4], Cathelijne M. J. Y. Tesink[3,4], Miriam Kos[3], and Peter Hagoort[1,3,5]

Abstract

■ When do listeners take into account who the speaker is? We asked people to listen to utterances whose content sometimes did not match inferences based on the identity of the speaker (e.g., ''If only I looked like Britney Spears'' in a male voice, or ''I have a large tattoo on my back'' spoken with an upper-class accent). Event-related brain responses revealed that the speaker's identity is taken into account as early as 200–300 msec after the beginning of a spoken word, and is processed by the same early interpretation mechanism that constructs sentence meaning based on just the words. This finding is difficult to reconcile with standard ''Gricean'' models of sentence interpretation in which comprehenders initially compute a local, context-independent meaning for the sentence (''semantics'') before working out what it really means given the wider communicative context and the particular speaker (''pragmatics''). Because the observed brain response hinges on voice-based and usually stereotype-dependent inferences about the speaker, it also shows that listeners rapidly classify speakers on the basis of their voices and bring the associated social stereotypes to bear on what is being said. According to our event-related potential results, language comprehension takes very rapid account of the social context, and the construction of meaning based on language alone cannot be separated from the social aspects of language use. The linguistic brain relates the message to the speaker immediately. ■

INTRODUCTION

We all use our knowledge of other people in making sense of what they say. For instance, we know that a 5-year-old is unlikely to say ''I'm going to quit smoking soon,'' and that it is really odd for a man to say ''I might be pregnant because I feel sick.'' If we know that somebody is a compulsive liar, or a politician running for election, we will interpret his or her words against that background. Thus, at some point during language comprehension, people combine the information that is represented in the contents of a sentence with the information they have about the speaker. We used event-related brain potentials (ERPs) to determine exactly when and how listeners relate what's being said to who is saying it.

In traditional linguistic theories about meaning (e.g., Grice, 1975), a distinction is often made between the context-free rule-based combination of fixed word meanings (''sentence meaning'') and the contributions made by the communicative context, such as who is speaking and what he or she might want (''utterance meaning'' or ''speaker meaning''). This way of partitioning meaning, which was reinforced by influential claims that basic sentence meaning could be derived from syntactic structure directly (Chomsky, 1957), has led to two rather separate subdisciplines in linguistics: *Semantics* deals with the rules and representations that govern sentence meaning, and *pragmatics* deals with the complexities introduced by the social, intentional aspects of communication. In psycholinguistics, this analysis of meaning has evolved into what we call the *standard two-step model of language interpretation*. In this model, listeners (and readers) first compute a local, context-independent meaning for the sentence, and only then work out what it really means given the wider communicative context and the particular speaker (Lattner & Friederici, 2003; Cutler & Clifton, 1999; Sperber & Wilson, 1995; Fodor, 1983; Grice, 1975). Mismatches between message and speaker would be detected in the second step only, in slow pragmatic computations that are different from the rapid semantic computations in which word meanings are combined (e.g., Lattner & Friederici, 2003).

From a design perspective, however, it would make sense for the linguistic brain to take the speaker into account right from the start. After all, human language has evolved to support social, interpersonal interaction. People use language to coordinate actions, transfer experience, regulate social status, and strengthen bonds, as well as to manipulate, intimidate, seduce and deceive. If these social aspects are so critical, why delay their use

[1]Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands, [2]University of Amsterdam, Amsterdam, Netherlands, [3]FC Donders Centre for Cognitive Neuroimaging, Nijmegen, Netherlands, [4]Radboud University Nijmegen, Nijmegen, Netherlands, [5]Nijmegen Institute for Cognition and Information, Nijmegen, Netherlands

in computing meaning? Also, recent linguistic research suggests that the computation of a context-free sentence meaning is, in fact, highly problematic, and that linguistic meaning is always colored by the pragmatics of the communicative exchange (Kempson, 2001; Perry, 1997; Clark, 1996). The meaning of so-called indexicals such as ''I'' and ''you,'' for example, inevitably depends on the communicative situation (e.g., Perry, 1997), and upon closer analysis, so does the meaning of apparently self-sufficient words such as ''garage'' or ''Kensington Gardens'' (Kempson, 2001; Clark, 1996). More generally, theorists have come to realize that linguistic communication is not so much about encoding, transferring, and decoding a message (with pragmatic context providing ''additional constraints''), but is an *intrinsically* contextualized and social activity in which speakers and listeners closely coordinate their joint behavior on the basis of what they know about each other (Clark, 1996). These analyses are at odds with the standard two-step model of interpretation. Instead, they suggest a one-step model in which knowledge about the speaker is brought to bear immediately by the same fast-acting brain system that combines the meanings of individual words into a larger whole.

Consistent with the latter, eye tracking studies on dialogue in structured conversational settings (usually a referential communication task) have shown that listeners are quickly sensitive to what they know about the perspective and other characteristics of the speaker (e.g., Trueswell & Tanenhaus, 2005; Hanna & Tanenhaus, 2004; Hanna, Tanenhaus, & Trueswell, 2003; Metzing & Brennan, 2003; see Barr & Keysar, 2006, for review). For instance, Metzing and Brennan (2003) found that listeners who, in conversation with a particular speaker, had converged upon a specific lexical description of an item in the scene (e.g., ''the shiny cylinder'') were delayed in looking at that item if the same speaker suddenly used a new description (e.g., ''the silver pipe''), but not if a *different* speaker used that new description. Also, Hanna and Tanenhaus (2004) showed that listeners who were asked to hand over something to the speaker were rapidly sensitive to whether the latter could have picked up the item himself or herself. Such eye tracking findings show that, in conversation, listeners rapidly relate the message to characteristics of the speaker. However, they cannot directly tell us whether the process (and underlying neural substrate) that merges the unfolding sentential message with information about the speaker is identical to, or different from, the process that combines word meanings.

Our research question is also related to a recent debate in the speech perception and sociophonetics literature. In traditional models of speech perception (see Nygaard, 2005, for discussion), characteristics of the speech signal, such as speaker identity or emotional tone of voice, are assumed to be processed separately from the recovery of ''sentence meaning.'' The general idea is that listeners initially strip away the acoustic variability associated with different speakers (''talker normalization'') to arrive at a standardized input representation from which they can subsequently recover the linguistic message. However, rather than rapidly disregarding voice-based cues to who the speaker is, listeners, in fact, use these cues in the earliest stages of speech signal processing (for reviews, see Johnson, 2005; Nygaard, 2005; Thomas, 2002). For example, listeners perceive vowels differently depending on whether a preceding stretch of filtered speech is shifted up or down in frequency to suggest a male or female speaker (Remez, Rubin, Nygaard, & Howell, 1987), and the perception of an ambiguous syllable-initial fricative is systematically modulated by whether the rest of the syllable is spoken by a man or a woman (Strand, 1999). Although these speech perception findings do not directly tell us when and how speaker identity merges with an unfolding sentential message, they demonstrate that listeners can extract and use information about the speaker from the acoustic signal very rapidly. Furthermore, merely showing the face of a male or female speaker also affects fricative and vowel perception (Johnson, Strand, & D'imperio, 1999; Strand, 1999), and when standard American English speech is accompanied by a picture of an Asian speaker, American listeners have a harder time making sense of the input (Rubin, 1992). This suggests that speaker identity is taken into account in the earliest stages of speech perception in a way that goes beyond a simple within-modality mechanism.

In our study, we examined the integration of message and speaker by means of scalp-recorded ERPs, a measure that allows us to selectively keep track of the various processes of language comprehension as they occur, with high temporal resolution. In the experiment, people listened to a pseudorandom mixture of sentences spoken by 21 different speakers. Some of these sentences contained a speaker inconsistency, a specific word at which the message began to mismatch probabilistic inferences about the speaker's sex, age, and social–economic status, as inferred from the speaker's voice. Examples: ''If only I looked like *Britney* Spears in her latest video'' in a male voice; ''Every evening I drink some *wine* before I go to sleep'' in a young child's voice; and ''I have a large *tattoo* on my back'' spoken in an upper-class accent. Other sentences contained a standard semantic anomaly, a specific word whose meaning did not fit the semantic context established by the preceding words, as in ''The earth revolves around the *trouble* in a year.''

Our research logic was based on a set of well-established facts about the *N400* component, a language-relevant negative deflection in the ERP peaking around 400–550 msec after spoken word onset, and largest at centro-posterior recording sites (see Kutas, Van Petten, & Kluender, 2006;

Kutas & Federmeier, 2000, for reviews). First, every content word in a sentence elicits an N400, but words that are semantically anomalous (e.g., ''trouble'' in the above example) elicit reliably larger N400s than words that are not (e.g., ''sun'', Kutas & Hillyard, 1980). Second, this differential N400 effect is associated with the analysis of *meaning* and is not elicited by syntactic, phonological, or spelling anomalies. Third, the N400 is not a simple anomaly detector. For example, N400 effects can also be elicited by equally coherent words that differ only in their predictability (e.g., Otten & Van Berkum, 2007; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; Hagoort & Brown, 1994; Kutas & Hillyard, 1984). Fourth, with spoken words, N400 effects begin to emerge after having heard only two or three phonemes, well before the word has ended (e.g., Van den Brink, Brown, & Hagoort, 2006; Van Berkum, Zwitserlood, Brown, & Hagoort, 2003; Van Petten, Coulson, Rubin, Plante, & Parks, 1999). According to these observations, the N400 effect elicited by semantic anomalies reflects some aspect of the normal early sense-making process during which every incoming word is related to the context established by the preceding words.

If inferences based on the voice of the speaker are recruited by the same early sense-making process that combines word meanings, and if these inferences become available rapidly enough as the sentence unfolds, then speaker inconsistencies and semantic anomalies should elicit the same ERP effect, an N400 effect. Note that under the one-step model, these two predicted N400 effects do not need to have the same size (an issue to which we return in the Discussion).[1] The critical prediction of this model is that because semantic information provided by the words in a sentence and voice-conveyed information about the identity of the speaker are handled by the same early sense-making process, semantic anomalies and speaker inconsistencies will generate the same type of ERP effect, an N400 effect, doing so in the typical latency range for N400 amplitude modulations. The two-step model of semantic interpretation makes a different prediction: If contextual information about the speaker is handled in a distinct second phase of interpretation, then speaker inconsistencies should elicit a delayed and possibly quite different ERP effect.

METHODS

Participants

Twenty-four neurologically unimpaired right-handed native speakers of Dutch, 12 men (19–22 years, mean age = 20.3 years) and 12 women (19–26 years, mean age = 21.9 years), were included. All participants gave informed consent in accordance with the Declaration of Helsinki.

Materials

We constructed 160 sentences with a lexical content that was fully consistent with one particular speaker, but substantially less consistent with another speaker. To increase variability, this set contained six types of speaker-inconsistent utterances: 40 were odd for a male speaker (''If only I looked like *Britney* Spears in her latest video''), 40 were odd for a female speaker (''On weekends I usually go *fishing* by the river''), 20 were odd for a young speaker (''Every evening I drink some *wine* before I go to sleep''), 20 were odd for an adult speaker (''I cannot sleep without my *teddy bear* in my arms''), 20 were odd for a speaker with an upper-class accent (''I have a large *tattoo* on my back''), and 20 were odd for a speaker with a lower-class accent (''Every month we go to an *opera* to have a night out''; see Appendix for more examples). Although some speaker inconsistencies were truly anomalous, in each of the six subtypes, the majority merely violated (Dutch) social stereotypes. Importantly, we designed the sentences such that the speaker-dependent inconsistency always emerged at a single critical word (italicized here, note that the English translation sometimes requires two words), and that the fragment before the critical word was fully compatible with either speaker (''In weekends I usually go. . .,'' ''I have a large. . .''). To give listeners some time to extract cues to the speaker's identity from the voice, at least three words preceded the critical word. Furthermore, to make sure our effects would not hinge on sentence-final wrap-up processes, critical words were never at the very end of the sentence. Between these two margins, we deliberately varied the position of the critical word.

We recorded all sentences with a consistent and inconsistent speaker (4 men and 4 women, 2 young children around age 6 and 8 years, and 2 adults, 2 speakers with a Dutch accent typically perceived as lower-class, and 2 with a Dutch accent typically perceived as upper-class), avoiding recordings in which the two contrasting speakers had used obviously different prosodic contours. Speaker-consistent and -inconsistent recordings were matched on duration of the critical words (speaker-consistent: mean = 520 msec, *SD* = 149 msec, range = 236–1023 msec; speaker-inconsistent: mean = 524 msec, *SD* = 140 msec, range = 212–921 msec), duration of the preceding sentence fragment (speaker-consistent: mean = 1595 msec, *SD* = 496 msec, range = 485–3367 msec; speaker-inconsistent: mean = 1626 msec, *SD* = 506 msec, range = 455–3261 msec), and time from critical word onset to sentence end (speaker-consistent: mean = 1585 msec, *SD* = 385 msec, range = 837–2593 msec); speaker-inconsistent: mean = 1595 msec, *SD* = 392 msec, range = 882–2812 msec).

To compare the ERP effect elicited by speaker inconsistencies to a standard N400 effect within the same group of subjects, we also included a supplementary set

of 96 items, 48 with a classic sentence-dependent semantic anomaly (e.g., ''Dutch trains are *sour* and blue''), and 48 with a semantically correct control (e.g., ''Dutch trains are *yellow* and blue''). Because our logic merely required a comparison of effect identity (notably polarity and scalp distribution) and did not hinge on a comparison of effect sizes, we made no attempt to match these and the speaker-dependent items on the degree of contextual fit in the respective critical and control variants. All 96 supplementary utterances were recorded with four neutral female speakers and one neutral male speaker, with the two variants of each item (e.g., ''Dutch trains are sour/yellow . . .'') always spoken by the same speaker. For purposes unrelated to the current issue, another 48 items contained a so-called world-dependent anomaly (e.g., ''Dutch trains are *white* and blue''; Hagoort, Hald, Bastiaansen, & Petersson, 2004), and a final 48 coherent items were true filler sentences, both again recorded with the abovementioned neutral speakers.

For each of six trial lists, we pseudorandomly mixed 80 speaker-inconsistent and 80 speaker-consistent utterances (proportionally balanced across the six speaker subtypes) with these 192 additional utterances, such that no participant heard the same sentence in more than one variant, each variant was heard by an equal number of participants, the longest consecutive sequence of trials of the same type was two, and such that each speaker produced an equal number of consistent and inconsistent utterances (with particular speakers in the speaker-consistency subdesign producing five of each kind). To reduce accidental variability in the data for repeated measures analysis, we divided the set of 180 speaker-relevant items such that for any subject, the 90 items in Condition A (e.g., consistent) were matched to the 90 items in the alternative condition (e.g., inconsistent) in terms of the acoustic duration and word frequency (on 3.7 million, *Corpus Spoken Dutch, Release 6*) of the critical word. The same was done for the 96 supplementary sentence-semantic items.

In a posttest conducted to validate the materials, another 12 men and 12 women listened to the same lists and were asked to rate on a 5-point scale ''how normal or strange you think it is to have the speaker say this particular thing'' (1 = completely normal, 5 = very strange). As expected, utterances that contained a speaker inconsistency were rated as less plausible (mean = 3.5, *SD* = 0.8, range = 1.5–5.0), than the corresponding speaker-consistent control utterances (mean = 1.6, *SD* = 0.4, range = 1.0–2.8). Furthermore, utterances that contained lexically dependent semantic anomalies were rated as highly implausible (mean = 4.6, *SD* = 0.3, range = 3.6–5.0), whereas the corresponding control utterances were considered to be acceptable (mean = 1.5, *SD* = 0.4, range = 1.0–4.3). Note that the average semantic anomaly was considered to be more problematic than the average speaker inconsistency. This is consistent with the fact that whereas the former were always anomalous (''The earth revolves around the trouble in a year,'' ''Dutch trains are sour and blue''), the latter often merely went against a social stereotype (such that Dutch women tend not to fish, or that people with Dutch upper-class accents are typically not expected to have a tattoo).

Procedure

After electrode application, participants sat in a sound-attenuating booth and listened to 352 sentences, spoken by 21 different people, and presented over audio speakers. We asked the participants to process each sentence for comprehension, and we did not impose any additional task. After a short practice, the trials were presented in five blocks of 10 min each, separated by rest periods. Each trial began with a fixation asterisk centered on the screen. After 1 sec, the spoken sentence was played from file. The asterisk remained on the screen until 1 sec after sentence offset, and was followed by a 3.6-sec intertrial interval. Participants were asked to avoid eye and other movements when the asterisk was visible, and to deliberately blink in the intertrial interval.

EEG Recording and Analysis

The electroencephalogram (EEG) was recorded from 28 cap-mounted silver–chloride electrodes (EasyCap), each referred to the left mastoid. Five electrodes were placed over the standard 10% system midline sites Fz, FCz, Cz, and Pz, and 11 pairs were placed over the standard lateral sites FP1/FP2, F7/F8, F3/F4, FC5/FC6, FC1/FC2, T7/T8, C3/C4, CP5/CP6, CP1/CP2, P7/P8, P3/P4, and O1/O2. Five additional electrodes, each also referred to the left mastoid, were used to aid in off-line signal processing: the right mastoid, two electrodes at the outer left and right canthi, and two electrodes above and below the left eye (converted off-line to bipolar horizontal and vertical EOG signals respectively). All electrode impedances were below 5 kΩ. Signals were recorded with a BrainAmps DC amplifier using a 200-Hz low-pass filter, a time constant of 10 sec (0.016 Hz), and a 500-Hz sampling frequency. After re-referencing the EEG signals to the mean of the left and right mastoid off-line, segments ranging from 500 msec before to 2000 msec after the acoustic onset of the critical word were baseline-corrected by subtracting the mean amplitude in the −150 to 0 msec prestimulus interval, and semiautomatically screened off-line for eye movements, muscle artifacts, electrode drifting, and amplifier blocking. Segments containing such artifacts were rejected (11.5%, with no asymmetry across conditions). The remaining EEG segments were averaged per participant and condition, and the associated mean amplitude values in specific latency ranges were submitted to repeated measures analyses of variance, using the Greenhouse–Geisser/Box's epsilon hat correction for

univariate F tests with more than one degree of freedom in the numerator (we report the original df). Effects were first evaluated in an overall analysis with all 28 electrodes, after which the topography was explored in a mean quadrant analysis involving the left anterior electrodes Fp1, F3, F7, FC1, FC5; the right anterior electrodes Fp2, F4, F8, FC2, FC6; the left posterior electrodes CP1, CP5, PO3, P7, O1; and the right posterior electrodes CP2, CP3, PO4, P8, O2 (defining an Anterior–posterior × Hemisphere design).

RESULTS

As can be seen in Figure 1, words at which the unfolding linguistic message began to mismatch voice-based inferences about the speaker elicited a small but clear N400 effect in brain potentials, with a classic maximum at electrode Pz, and a classic time course between 200 and 700 msec after acoustic word onset. The speaker-dependent N400 effect was reliable overall [e.g., $F(1, 23) = 5.47$, $MSE = 10.61$, $p = .028$ across all electrodes in the 200–700 msec latency range], and varied in size across the 28 electrodes [Consistency × Electrode interaction: $F(1, 27) = 2.83$, $MSE = 2.24$, $p = .028$]. A quadrant analysis showed that the effect was larger over posterior than anterior regions of the scalp [$F(1, 23) = 5.21$, $MSE = 0.52$, $p = .032$]. However, in the same analysis, effect size did not reliably depend on hemisphere ($F < 1$, $p = .617$), nor on hemisphere and anteriority considered together ($F < 1$, $p = .781$). Follow-up analysis revealed a reliable N400 effect across all 11 posterior electrodes [−0.66 μV difference, $F(1, 23) = 14.66$, $MSE = 3.88$, $p = .001$], but not across all 12 anterior electrodes ($F < 1$, $p = .392$).

Because the speaker-dependent N400 effect is small and predominantly posterior, we examined specific latency ranges at the 11 posterior electrodes only. First, a reliable speaker inconsistency effect emerged in the 300–500 msec latency range, the standard latency range used for quantifying N400 effects [−0.56 μV difference, $F(1, 23) = 5.62$, $MSE = 7.28$, $p = .027$]. Second, reliable effects of speaker inconsistency were also found for the immediately following 500–700 msec [−0.87 μV difference, $F(1, 23) = 20.45$, $MSE = 4.88$, $p < .001$] and the

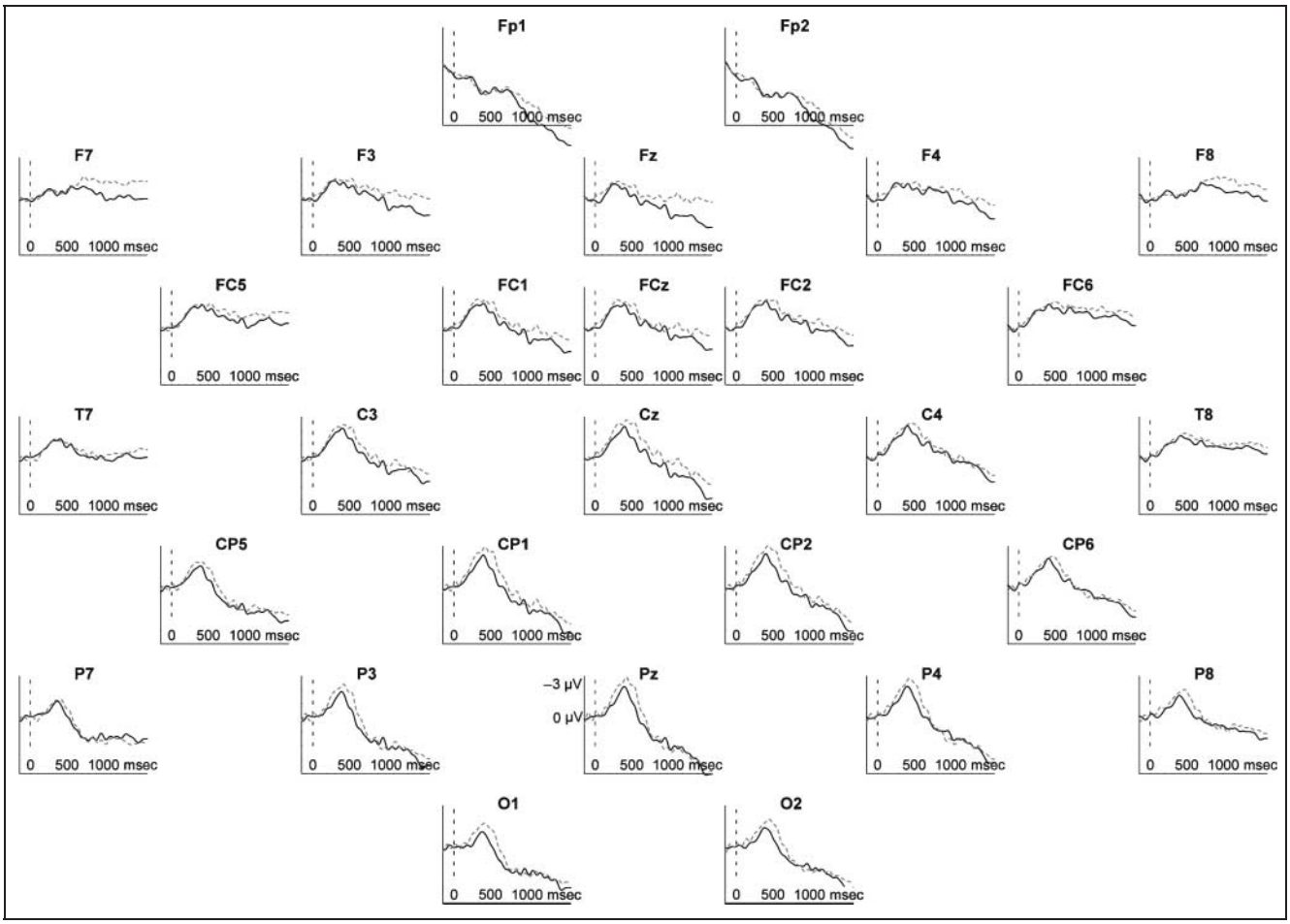

Figure 1. Speaker inconsistency effect. Grand-average ERPs to words whose meaning did (solid) or did not (dotted) easily fit voice-based inferences about the speaker, pooled across speaker dimensions. In this and all following figures, negative voltage is up, and acoustic onset of the critical word is at 0 msec.

immediately preceding 200–300 msec [−0.43 μV difference, $F(1, 23) = 4.80$, $MSE = 5.12$, $p = .039$]. No reliable effect of speaker inconsistency emerged before this [e.g., 100–200 msec: $F(1, 23) = 1.39$, $MSE = 7.11$, $p = .250$].

As shown in Figure 2, words that were anomalous in the local sentence-semantic context elicited a very large N400 effect [$F(1, 23) = 44.61$, $MSE = 30.86$, $p < .001$, across all electrodes in the 200–700 msec latency range]. Like the speaker-dependent N400 effect, the semantic anomaly effect first emerged in the 200–300 msec range [$F(1, 23) = 7.38$, $MSE = 13.72$, $p = .012$ across all 11 posterior electrodes], with no reliable effect before this [100–200 msec: $F(1, 23) = 2.27$, $MSE = 6.87$, $p = .145$]. With a mean posterior effect size of −2.71 μV in the 200–700 msec range, this ''classic'' N400 effect is approximately four times as large as the speaker-dependent N400 effect [Inconsistency × Sentence type: $F(1, 23) = 31.48$, $MSE = 8.88$, $p < .001$]. But when the size difference of the two effects in this latency range is adjusted for by differentially scaling the scalp topography plots by a factor four, we find virtually identical scalp distributions, shown together in Figure 3. This was confirmed by an additional analysis of variance on scaled data. After we divided the electrode-specific amplitude values for each participant and condition by the mean for that participant and condition, the scalp distribution of the two effects did not differ significantly [Inconsistency × Sentence type × Electrode: $F < 1$, $p = .835$].

Apart from an early N400 effect, speaker inconsistencies and semantic anomalies also elicited a late anterior negative shift [across all 12 anterior electrodes and 1000–1500 msec, speaker inconsistency effect −0.80 μV: $F(1, 23) = 9.91$, $MSE = 9.23$, $p = .005$; semantic anomaly effect −1.26 μV: $F(1, 23) = 5.83$, $MSE = 39.07$, $p = .024$; interaction $F < 1$, $p = .392$]. These late anterior shifts possibly reflect additional mid-sentence inferencing triggered by a conceptual problem (see Van Berkum et al., 2003, for another example), and might as such be related to the anterior negative shift or *Nref* effect elicited by referential ambiguities (see Van Berkum, Koornneef, Otten, & Nieuwland, 2007, for review). However, because some of our critical sentences end as early as 750 msec after critical word onset, the late shift may also reflect sentence-final wrap-up processes elicited by the ends of problematic sentences.

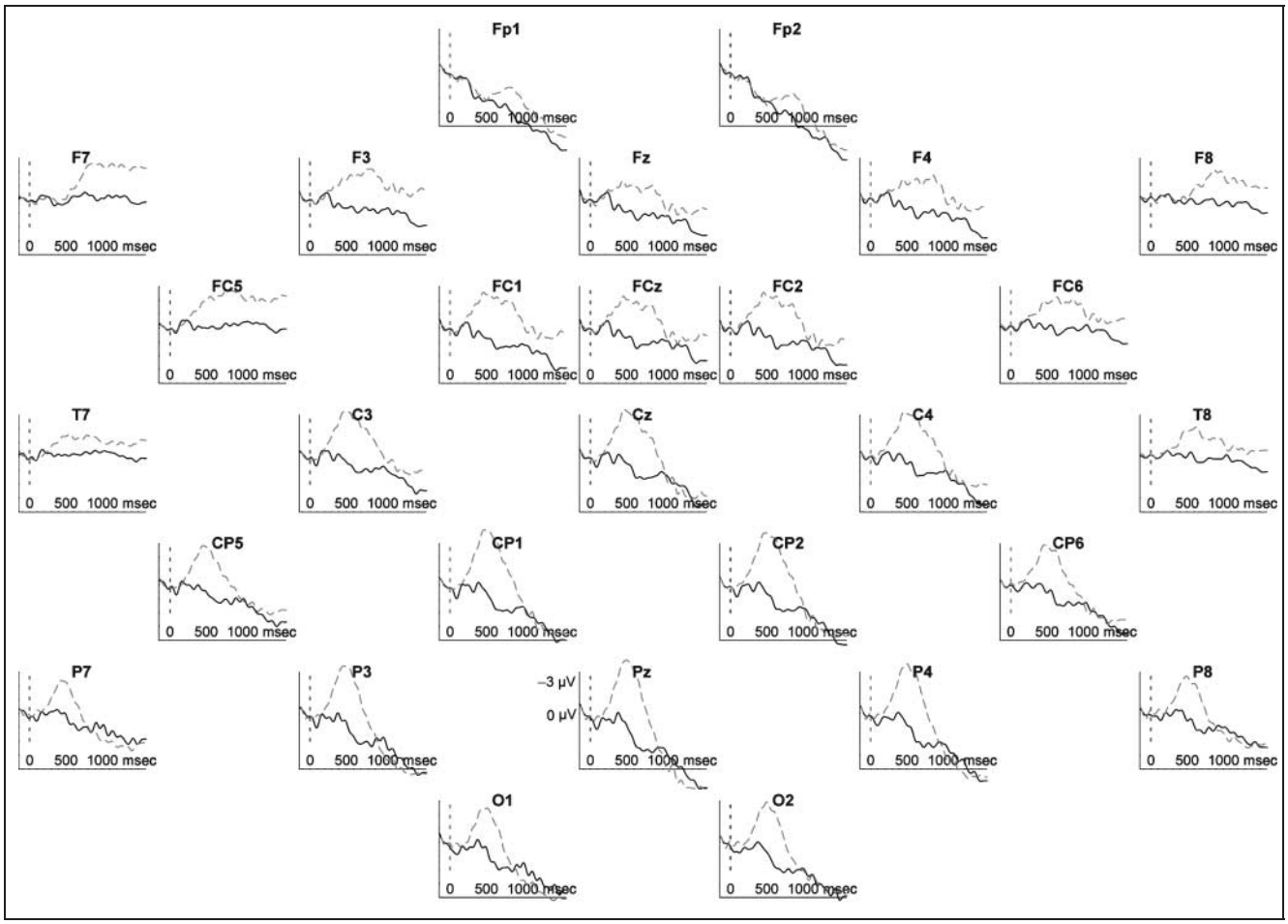

Figure 2. Semantic anomaly effect. Grand-average ERPs to words whose meaning was coherent (solid) or anomalous (dashed) in the prior lexical–sentential context.

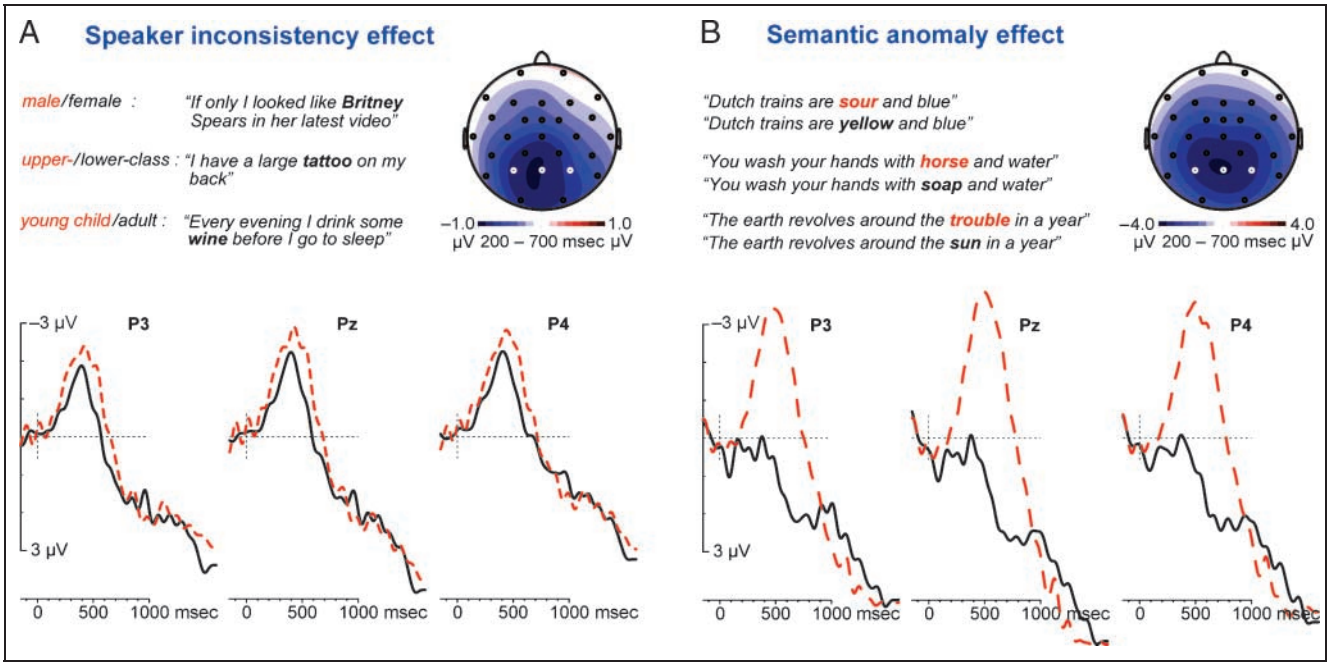

Figure 3. Speaker inconsistency effect and semantic anomaly effect at three posterior electrodes, with identical (but differentially scaled) scalp topographies. Although the effects differ in size by a factor four, speaker inconsistencies and semantic anomalies both elicited the classic N400 effect that is known to reflect early sense-making processes in language comprehension.

In a post hoc analysis, we examined whether the speaker inconsistency effect depended on voice dimension, by comparing the ERP effects elicited by male/female-based inconsistencies to those elicited by age and upper/lower-class accents (pooled together to obtain a reasonable signal-to-noise ratio). The two N400 effects did not reliably differ in size [$F(1, 22) = 1.23$, $MSE = 13.89$, $p = .280$, over all posterior electrodes in the 200–700 msec latency range]. However, inconsistencies that hinged on male or female voices elicited an additional late positivity [0.83 μV at electrode Pz in the 700–1200 msec latency range, $F(1, 22) = 5.79$, $MSE = 1.42$, $p = .025$; 0.57 μV over all posterior electrodes, $F(1, 22) = 3.98$, $MSE = 10.87$, $p = .058$], which was not observed for the other two dimensions [Inconsistency by Voice dimension: $F(1, 22) = 6.73$, $MSE = 16.10$, $p = .017$]. We briefly return to this effect below.

DISCUSSION

According to our ERP results, the brain integrates message and speaker very rapidly, within some 200–300 msec after the acoustic onset of a relevant word. Also, speaker inconsistencies elicited the same *type* of brain response as semantic anomalies, an N400 effect. That is, voice-inferred information about the speaker is taken into account by the same early language interpretation mechanisms that construct ''sentence-internal'' meaning based on just the words. Our findings therefore demonstrate that, as far as the brain is concerned, linguistic meaning depends on the pragmatics of the communicative situation right from the start. Other evidence from the N400 already indicated that words are immediately related to a prior narrative discourse (e.g., Nieuwland & Van Berkum, 2006; Van Berkum et al., 2003; Van Berkum, Hagoort, & Brown, 1999; St. George, Mannes, & Hoffman, 1994), and to one's knowledge of the world (Hagoort et al., 2004). However, by revealing an equally immediate impact of what listeners infer about the *speaker*, the present results add a distinctly social dimension to the mechanisms of on-line language interpretation. Language users very rapidly model the speaker to help determine what is being said. This makes sense, as language evolved in face-to-face social interaction and, importantly, requires close coordination among interlocutors (Clark, 1996).[2]

The average critical word in our study lasted 522 msec, and only 9 out of 320 words were shorter than 300 msec. This suggests that listeners already relate what is being said to who is saying it as the relevant spoken word unfolds. With a preceding sentence context that lasted, on average, 1.6 sec and varied between 0.5 and 3.4 sec, it seems safe to assume that, in most cases, our listeners had some idea already as to the plausible sex, age, and social stratum of the speaker before the critical word began. However, what is interesting is that the—usually stereotype-mediated—implications of these voice-based identifications already kick in before the critical word has been fully heard. We already knew from other work that sentence- and discourse-dependent N400 effects begin to emerge after having heard only two or three phonemes, well before the word has ended (e.g., Van

den Brink et al., 2006; Van Berkum et al., 2003; Van Petten et al., 1999). What we see now is that voice-dependent inferences about who is speaking are brought to bear on comprehension in the same early latency range.

Our findings converge with eye movement evidence for the rapid use of speaker-related information during comprehension in referential communication tasks (e.g., Trueswell & Tanenhaus, 2005; Hanna & Tanenhaus, 2004; Hanna et al., 2003; Metzing & Brennan, 2003). However, our voice-dependent ERP results go beyond the extant findings in several ways. First, in referential communication paradigms, the listener can draw inferences about speakers that they see and interact with, but the listeners in our experiment only had a voice to go on. The relevant features that together define voice quality have turned out to be quite difficult to pin down systematically (Kreiman, Vanlancker-Sidtis, & Gerratt, 2005), and can include such things as fundamental frequency, vowel formant frequencies, and timing. Entirely consistent with everyday experience, our ERP findings show that listeners can rapidly extract and use these features to classify the speaker along socially important dimensions.

Second, in the abovementioned referential communication studies, the range of speaker characteristics relevant to interpretation is usually highly constrained, typically boiling down to whether the speaker can or cannot see (hence, refer to) a particular object in the scene. In our study, the situation was much more open-ended, for as the utterance unfolded, *any* (usually stereotype-mediated) property of the speaker could turn out to be relevant. Of course, we used only three speaker dimensions to realize the speaker inconsistencies, and participants may well have caught on to this. But what makes ''On weekends I usually go fishing by the river'' in a women's voice atypical is not just the sex of the speaker but the intuition that (Dutch) women tend not to fish, and what makes ''I have a large tattoo on my back'' in a stereotypically upper-class voice odd is not just the social class identification but the intuition that (Dutch) upper-class behavior tends not to include getting a tattoo. Because the initial part of the speaker-relevant sentences had to fit *either* speaker and, therefore, usually had a relatively shallow, nonpredictive content (e.g., ''On weekends I usually go. . .,'' ''I have a large. . .''), the sentential context provided no information as to which plausible characteristic of the speaker would be relevant before the critical word (''fishing,'' ''tattoo'') came along. This suggests that relevant inferences about the speaker can be drawn and brought to bear on language processing extremely rapidly, even in very open-ended situations.

A final and more general difference is that, unlike eye movements, brain potentials provide clear cues to the *identity* of the processes involved, with 0-msec delay. They therefore allow for stronger inferences about whether two sources of information are recruited by the same process, the issue under investigation here. The equivalence of our two critical ERP effects, in polarity, scalp distribution, and latency range, are most parsimoniously interpreted as generated by a common (set of) neuronal generator(s). As such, this equivalence speaks against the classic two-step model of language interpretation, and provides unique support for constraint-based models of comprehension (Jackendoff, 2002; Tanenhaus & Trueswell, 1995; MacDonald, Pearlmutter, & Seidenberg, 1994). In the latter, constraints that are sufficiently salient and relevant, no matter what their source, can all simultaneously help determine interpretation, in a unified computational system, and without the principled delays postulated by the standard two-step model. This is exactly what our ERP data suggest. It is also consistent with recent fMRI evidence that speaker inconsistencies and semantic anomalies engage the same brain area (BA 45/47 in the left inferior prefrontal cortex; Tesink et al., 2007).

Why is the Speaker Effect So Small?

Although we did not control the differential degree of fit across speaker- and semantics-relevant sentences, one might find it surprising that the speaker-dependent N400 effect is so much smaller than the semantic anomaly effect (see Figure 3). We believe this result reflects at least two specific incidental properties of our stimulus materials, and should thus not be taken as evidence that constraints derived from who the speaker is *necessarily* matter less to initial sense-making processes than constraints derived from the sentence context. First, whereas the coherent control words in our semantic anomaly sentences were generally rather predictable (''The earth revolves around the *sun*,'' predictability is often used to help maximize N400 effects), the coherent control words in our speaker-consistent sentences were not predictable at all. This is because the preceding fragment (''On weekends I usually go. . .,'' ''I have a large. . .'') had to fit either speaker. Because less predictable words elicit larger N400 deflections (e.g., Otten & Van Berkum, 2007; Hagoort & Brown, 1994; Kutas & Hillyard, 1984), this explains the large difference in ''baseline'' N400 deflections elicited by the two types of control words (black solid lines in Figure 3). Second, whereas all of the semantic anomaly items were severely anomalous, most of our speaker inconsistencies merely went against a defeasible social stereotype, for instance, that Dutch women *tend* not to fish, or that people with upper-class accents *usually* do not have a large tattoo. This acceptability difference, confirmed by an independent rating study on the same materials (see Methods), can help explain why the N400 to semantic anomalies is somewhat larger than the N400 to speaker inconsistencies (red dotted/dashed lines in Figure 3).

It not unlikely that the difference in N400 effect size observed here may, in part, also reflect something

interesting about the relative degree of constraint provided by words and voices. In the context of a study on emotional prosody and word sense disambiguation, Nygaard and Lunders (2002; see also Nygaard, 2005) have speculated that, whereas voice cues and sentential context may constrain interpretation in the same way, that is, via the same mechanism, the constraints provided by tone of voice or prosody might, on average, be somewhat weaker due to the fact that these are usually more probabilistic than lexical–semantic cues. This argument also holds for our study, as voice-based cues on speaker sex, age, and social stratum are inherently probabilistic as well (e.g., some women have a low voice), and may, as such, have provided somewhat weaker average constraints on interpretation than the lexical–semantic contexts. In fact, we cannot exclude that, faced with conflicting constraints, listeners in our experiment may, on particular occasions, actually have doubted whether they ''got the speaker right.'' By comparing the impact of a voice cue (e.g., a female speaker) to an equivalent lexical–semantic sentence context (e.g., ''The woman said that...''), it should be possible to more systematically assess the relative force of lexical and voice-conveyed cues to interpretation.

New Questions

Our findings raise several new questions. One is to what extent the results depend on the self-referential pronouns ''I,'' ''mine,'' ''we,'' or ''our'' in our critical sentences. We used these pronouns deliberately to maximize the probability of a relevant effect. However, in the one-step model, conceptual speaker inconsistencies that do *not* depend on self-reference, such as a 5-year-old child mentioning the laws of motion, should also generate an N400 effect. We therefore predict that although the presence of self-referential pronouns may enhance the size of a speaker-dependent N400 effect, such pronouns are not critical.

A second question is whether the speaker dimension matters. Our study did not reveal a reliable difference between the N400 effects elicited by male/female speaker inconsistencies and by those involving age and upper/lower-class accents (pooled together to obtain a reasonable signal-to-noise ratio). Thus, there are no grounds to assume that the specific speaker dimension matters to early sense-making processes. However, only male/female speaker inconsistencies elicited an *additional* late posterior positivity. It is as yet unclear whether this additional ERP effect reflects something principled about how these particular speaker inconsistencies are dealt with after initial detection, or instead reflects specific incidental differences between the item subsets involved (e.g., only male/female speaker inconsistencies were sometimes biologically impossible, as ''I don't like having my period when I'm on vacation'' in a male voice). We note that in an earlier ERP study with sex-dependent speaker inconsistencies only, Lattner and Friederici (2003) also observed a posterior positivity in this latency range. Whether these effects are related (see below for important differences between the two experiments), and whether there is a relation to other meaning-induced late positivities (see Kuperberg, 2007, for review) remains to be established.

In contrast to our study, the Lattner and Friederici (2003) experiment did not reveal an early N400 effect to sex-dependent speaker inconsistencies. We suspect that this difference may have come about because of the specific design of the latter experiment. First, participants heard a mixture of 140 speaker-consistent and 140 speaker-inconsistent critical short sentences without fillers, with inconsistencies that invariably depended on the male/female voice contrast and always arose at the sentence-final word. This combination of features may not only have helped participants to discover the critical manipulation but may also have prompted them to approach the materials in a different way than they would in natural language comprehension. Furthermore, each of the male or female speakers said something atypical 17 to 18 times in any one experimental session. This may have allowed listeners to become acquainted with the speakers as specific individuals who do not fit the gender stereotype, a type of learning that might reduce or even fully eliminate inconsistency effects that depend on stereotypical expectations. The latter scenario, although in need of further testing, is consistent with the evidence for long-lasting implicit memory traces for specific voices (Goldinger, 1996).

Finally, one may ask whether, in a sufficiently sensitive and nontransparent study, inconsistencies between speaker and linguistic utterance should always modulate the N400. We see no reason why this would be the case. A word whose syntactic category disconfirms the listener's expectation about the typical syntactic structures used by a very young speaker, for example, may well elicit a P600 effect, that is, the ERP effect typically associated with violations of sentence- or text-induced syntactic expectations (e.g., Van Berkum, Brown, & Hagoort, 1999; Hagoort, Brown, & Groothusen, 1993; Osterhout & Holcomb, 1992). Likewise, if the referents that a listener considers for a referring expression (''it,'' ''the girl,'' ''the cake mix'') immediately depend on the perspective and other characteristics of the speaker (e.g., Trueswell & Tanenhaus, 2005; Hanna & Tanenhaus, 2004; Hanna et al., 2003), knowledge about the latter should be able to affect the specific ERP indices and neuronal systems associated with referential ambiguity and referential failure (e.g., Nieuwland, Petersson, & Van Berkum, 2007; Van Berkum et al., 2007). Thus, the early neuronal effects of what listeners know about the speaker should depend on the specific level of linguistic analysis (semantic, syntactic, referential; cf. Jackendoff, 2002) that this knowledge is *most relevant to*.

Conclusion

The ERP data reported here show that listeners use what they know or infer about the speaker from his or her voice in the earliest stages of meaning construction. This suggests that, to the brain of the language user, there is no context-free meaning. Instead, according to these findings, sentence interpretation is an intrinsically contextualized social activity (Kempson, 2001; Clark, 1996). Our results are difficult to reconcile with two-step ''Gricean'' models of sentence interpretation based on the classic distinction between semantics and pragmatics. As such, they could also be taken to query the nature of the distinction itself. Of course, the implication is not that listeners cannot tell the difference between a message and a speaker, and that everything blurs into an undifferentiated whole—the fact that we can perceive and reflect upon the conflict in a male voice uttering ''I think I am pregnant'' reveals that we can recover the lexically coded part of the message regardless of who the speaker is. However, what the brain's electrophysiology allows us to see is that voice-based inferences about the identity of the speaker and information encoded in the meaning of spoken words jointly constrain the same early sense-making process, without a principled delay in the use of speaker-related information. The linguistic brain is not just combining words in a context-free semantic universe confined in a single person's skull. It immediately cares about other people.

APPENDIX

Additional speaker (in)consistency examples for each speaker type, translated from Dutch, with numbers proportional to how many of each type featured in a list. Critical words that rendered the message inconsistent with the indicated voice are in italics and, due to translation, sometimes distributed across two words.

Inconsistent with male voice:

1. I recently had a check-up at the *gynecologist* in the hospital.
2. When I watch TV I often *cry* during a good movie.
3. Before I leave I always check whether my *make-up* is still OK.
4. I bought the same *sewing machine* as Ella did.
5. My job in *kindergarten* is really perfect for me.
6. My favorite colors are *pink* and apple green.

Inconsistent with female voice:

1. I broke my ankle playing *soccer* with friends.
2. As we moved house I carried the *washing machine* up the stairs.
3. The day starts best when I drive my *tractor* into the field.
4. At work I always have to wear a *tie* with the Shell logo.
5. For my birthday I got a *hammer drill* from my best friend.
6. Just before the counter I dropped my *aftershave* on the floor.

Inconsistent with young child's voice:

1. I really love *olives* with garlic.
2. Last year I got *married* in a beautiful castle.
3. I always read the *newspaper* before I leave.

Inconsistent with adult voice:

1. Twice a week I have *swimming lessons* in a very big pool.
2. My favorite book is the *fairy tale* Sleeping Beauty.
3. On the beach I made *sandcastles* by the sea.

Inconsistent with upper-class accent:

1. In the evening I often go to a *burger joint* for a hamburger.
2. Because of my job I spend a lot of time in the company *truck* on the road.
3. I take my two *pitbulls* anywhere I go.

Inconsistent with lower-class accent:

1. In my garage I have a *Jaguar* with leather upholstery.
2. My wife works as a *judge* in criminal law.
3. On Sundays I always play *golf* with some friends.

Acknowledgments

Partly supported by an NWO *Innovation Impulse* grant to J. V. B. We thank Petra van Alphen, Herb Clark, José Kerstholt, Oliver Müller, and two anonymous reviewers for their help.

Reprint requests should be sent to Jos J. A. Van Berkum, Max Planck Institute for Psycholinguistics, PO Box 310, 6500 AH Nijmegen, Netherlands, or via e-mail: jos.vanberkum@mpi.nl, Web: www.josvanberkum.nl.

Notes

1. In ERP research, arguments on whether different manipulations engage the same common process typically focus on polarity, scalp distribution, and coarse timing of the respective ERP effects. Differences in effect size as well as fine differences in timing are usually discarded as irrelevant to this issue. Differences in N400 effect size across studies and manipulations (input modality, anomaly vs. unexpectedness, etc.), for example, are usually taken to reflect incidental properties of the stimulus materials and/or experimental setting, rather than as indications that functionally different processes are involved (cf. Kutas & Federmeier, 2000, Figure 1).
2. Recent evidence suggests that mismatches between emotional prosody and emotional word meaning (e.g., ''failure'') also modulate the N400 (Schirmer & Kotz, 2006; Schirmer, Kotz, & Friederici, 2002, 2005). Although obtained in a somewhat unnatural language comprehension paradigm, in which the critical word targets were supplied in written form for a lexical decision and separated from a prosody-bearing prime sentence, this result is consistent with our findings and the perspective behind it: Listeners model the speaker to rapidly make sense of linguistic input.

REFERENCES

Barr, D. J., & Keysar, B. (2006). Perspective-taking and the coordination of meaning in language use. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 901–938). Amsterdam: Elsevier.

Chomsky, N. (1957). *Syntactic structures*. Den Haag: Mouton.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University 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: Oxford University Press.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge: MIT Press.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1166–1183.

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.

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., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes, 8,* 439–483.

Hagoort, P., Hald, L., Bastiaansen, M. C. M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304,* 438–440.

Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science, 28,* 105–115.

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language, 49,* 43–61.

Jackendoff, R. (2002). *Foundations of language*. New York: Oxford University Press.

Johnson, K. A. (2005). Speaker normalization in speech perception. In D. B. Pisoni & R. E. Remez (Eds.), *Handbook of speech perception* (pp. 363–389). Oxford: Blackwell.

Johnson, K. A., Strand, E. A., & D'imperio, M. (1999). Auditory–visual integration of talker gender in vowel perception. *Journal of Phonetics, 24,* 359–384.

Kempson, R. (2001). Pragmatics: Language and communication. In M. Aronoff & J. Rees-Miller (Eds.), *Handbook of linguistics*. Malden, MA: Blackwell.

Kreiman, J., Vanlancker-Sidtis, D., & Gerratt, B. R. (2005). Perception of voice quality. In D. B. Pisoni & R. E. Remez (Eds.), *Handbook of speech perception* (pp. 338–362). Oxford: Blackwell.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research, 1146,* 23–49.

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. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207,* 203–205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307,* 161–163.

Kutas, M., Van Petten, C., & Kluender, R. (2006). Psycholinguistics electrified II: 1994–2005. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 659–724). New York: Elsevier.

Lattner, S., & Friederici, A. D. (2003). Talker's voice and gender stereotype in human auditory sentence processing—evidence from event-related brain potentials. *Neuroscience Letters, 339,* 191–194.

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

Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects in the comprehension of referring expressions. *Journal of Memory and Language, 49,* 201–213.

Nieuwland, M. S., Petersson, K. M., & Van Berkum, J. J. A. (2007). On sense and reference: Examining the functional neuroanatomy of referential processing. *Neuroimage, 37,* 993–1004.

Nieuwland, M. S., & Van Berkum, J. J. A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience, 18,* 1098–1111.

Nygaard, L. C. (2005). Perceptual integration of linguistic and non-linguistic properties of speech. In D. B. Pisoni & R. E. Remez (Eds.), *Handbook of speech perception* (pp. 390–413). Oxford: Blackwell.

Nygaard, L. C., & Lunders, E. R. (2002). Resolution of lexical ambiguity by emotional tone of voice. *Memory & Cognition, 30,* 583–593.

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. (2007). What makes a discourse constraining? Comparing the effects of discourse message and scenario fit on the discourse-dependent N400 effect. *Brain Research, 1153,* 166–177.

Perry, J. (1997). Indexicals and demonstratives. In C. Wright & R. Hale (Eds.), *Companion to the philosophy of language* (pp. 586–612). Oxford: Blackwell.

Remez, R. E., Rubin, P. E., Nygaard, L. C., & Howell, W. A. (1987). Perceptual normalization of vowels produced by sinusoidal voices. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 40–61.

Rubin, D. (1992). Nonlanguage factors affecting undergraduates' judgements of non-native English speaking teaching assistants. *Research in Higher Education, 33,* 511–531.

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences, 10,* 24–30.

Schirmer, A., Kotz, S. A., & Friederici, A. D. (2002). Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research, 14,* 228–233.

Schirmer, A., Kotz, S. A., & Friederici, A. D. (2005). On the role of attention for the processing of emotions in speech: Sex differences revisited. *Cognitive Brain Research, 24,* 442–452.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition*. Oxford: Blackwell.

St. George, M., Mannes, S., & Hoffman, J. E. (1994). Global semantic expectancy and language comprehension. *Journal of Cognitive Neuroscience, 6,* 70–83.

Strand, E. A. (1999). Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology, 18,* 86–99.

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: Academic Press.

Tesink, C., Hagoort, P., Petersson, K., Van Berkum, J. J. A., Van der Gaag, R., Kan, C., et al. (2007). *Pragmatic language*

comprehension in adults with ASD: An fMRI study. Presented at the International Meeting for Autism Research (IMFAR), Seattle, May 3–5.
Thomas, E. R. (2002). Sociophonetic applications of speech perception experiments. *American Speech, 77,* 115–147.
Trueswell, J., & Tanenhaus, M. (Eds.) (2005). *Approaches to studying world-situated language use: Bridging the language-as-product and language-action traditions*. Cambridge: MIT Press.
Van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999). 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., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 443–467.
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., Koornneef, A. W., Otten, M., & Nieuwland, M. S. (2007). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research, 1146,* 158–171.
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 den Brink, D., Brown, C. M., & Hagoort, P. (2006). The cascaded nature of lexical selection and integration in auditory sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 364–372.
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.