Integration of Word Meaning and World Knowledge in Language Comprehension

Peter Hagoort,1,2,3* Lea Hald,1 Marcel Bastiaansen,1 Karl Magnus Petersson1

1F.C. Donders Centre for Cognitive Neuroimaging, 2Nijmegen Institute for Cognition and Information, University of Nijmegen, Nijmegen, The Netherlands. 3Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

*To whom correspondence should be addressed. E-mail: peter.hagoort@fcdonders.kun.nl

Although the sentences that we hear or read have meaning, this does not necessarily mean that they are also true. Relatively little is known about the critical brain structures for, and the relative time course of, establishing the meaning and truth of linguistic expressions. Here we present EEG data that show the rapid parallel integration of both semantic and world knowledge during the interpretation of a sentence. fMRI data revealed that the left inferior prefrontal cortex is involved in the integration of both meaning and world knowledge. Finally, oscillatory brain responses indicate that the brain keeps a record of what makes a sentence hard to interpret.

Language is used, among other things, to exchange information about the world. This entails that during on-line comprehension, the meaning of a phrase or sentence is derived and, in many cases, its truth is verified. To be able to do so, usually information about the words of a language and about the facts of the world need to be retrieved from memory.

At least since Frege (1, 2), theories of meaning make a distinction between the semantics of an expression, and its truth-value in relation to our mental representation of the state of affairs in the world (3). For instance, the sentence “The present queen of England is divorced” has a coherent semantic interpretation, but contains a proposition that is false in the light of our knowledge in memory that her majesty is married to Prince Phillip. The situation is different for the sentence “The favorite palace of the present queen of England is divorced.” Under default interpretation conditions, this sentence has no coherent semantic interpretation, since the predicate is-divorced requires an animate argument. This sentence mismatches with our representation of the world in memory, because the descriptive features of the purported state of affairs are inherently in conflict. The difference between these two sentences alludes to the distinction that can be made between facts of the world, and facts of the words of our language, including their meaning. Although theories of semantic memory usually do not make this distinction (4), accounts of on-line language processing often do, and distinguish between the retrieval and usage of world knowledge and of knowledge of word meaning.

In relation to the distinction between facts of the world and facts of the words of one’s language, some aspects of word meaning might be characterized as linguistic in nature, while other aspects relate to world knowledge. In linguistic theory, the latter is referred to as the domain of pragmatics, and the former as the domain of semantics. Based on this distinction between semantics and pragmatics (5–7), the semantic interpretation of a sentence is thought to be separate from and to precede the integration of pragmatic or world knowledge information (8). However, a number of researchers (e.g., 3, 9) have pointed out that the distinction between linguistic meaning and world knowledge is problematic, since many words are polysemous, and their meaning can only be fully established by invoking world knowledge (10).

We decided to contribute to settling this issue by providing neurophysiological evidence on the integration of semantic and world knowledge information. The underlying idea is that if a principled distinction can be made between linguistic meaning and world knowledge, concomitant processing differences should be obtained during the interpretation of a sentence.

This study presents EEG and fMRI data that speak to this issue. While participants’ brain activity was recorded, they read three versions of sentences such as: “The Dutch trains are yellow/white/sour and very crowded.” (the critical words are in italics). It is a well-known fact among Dutch people that Dutch trains are yellow, and therefore the first version of this sentence is correctly understood as true. However, the linguistic meaning of the alternative color term white applies equally well to trains as the predicate yellow. It is world knowledge about trains in Holland that makes the second version of this sentence false. This is different for the third version. This version contains a violation of semantic constraints. The core meaning of sour is related to taste and food. Under standard interpretation conditions a predicate requires an argument whose semantic features match that of its predicate. For vehicles such as trains, this is clearly not the case, since semantic features related to taste and food do not
apply to the materials that trains are made of. One could thus argue that for semantic-internal reasons the third sentence is false or incoherent (3, 11). It is our knowledge about the words of our language and their linguistic meaning that pose a problem for the interpretation of the third version of this sentence.

The increased interpretation load of semantic and world knowledge violations is assumed to have an effect on electrophysiological brain activity and on the hemodynamic response. If semantic interpretation precedes verification against world knowledge, the effects of the semantic violations should be earlier and might invoke other brain areas than the effects of the world knowledge violations. Based on EEG recordings from 29 electrode sites (12, 13), event-related brain potentials (ERPs) were computed, time-locked to the onset of the critical words that embodied the semantic violation, the world knowledge violation, and their correct counterpart, respectively. ERPs reflect the summation of the postsynaptic potentials of a large ensemble of synchronously active neurons. They provide a sampling of the brain’s electrical activity with a very high temporal resolution. In the present study, we focused on one particular ERP effect, referred to as the N400. The amplitude of this negative-going ERP between roughly 250 and 550 ms, with a maximum at about 400 ms, is influenced by the processing of semantic information (14). The easier the match between the lexical-semantics of a particular content word and the semantic specification of the context, the more reduced the N400 amplitude will be. The N400 is known to be very sensitive to semantic integration processes (15, 16).

As expected, the classic N400 effect was obtained for the semantic violations. For the world knowledge violations we also observed a clear N400 effect. Crucially, this effect was identical in onset and peak latency, and very similar in amplitude and topographic distribution to the semantic N400 effect (Fig. 1). This finding is strong empirical evidence that lexical-semantic knowledge and general world knowledge are both integrated in the same time-frame during sentence interpretation, starting at about 300 ms after word onset. In addition, we used the EEG data to investigate oscillatory brain activity in a wide frequency range (1–70 Hz) in relation to the semantic and world knowledge violations. Amplitude increases of EEG oscillations in specific frequency bands, such as theta (4–7 Hz) and gamma (approximately 30–70 Hz) that are induced by a cognitive event, are thought to reflect the dynamic recruitment of the relevant neuronal networks engaged in cognitive processing (17). A wavelet-based time-frequency representation of EEG power changes (Fig. 2) revealed a clear gamma peak for the world knowledge violation that was not seen for the semantic violation (18). In contrast, relative to the other conditions the semantic violation resulted in an increase in power in the theta frequency range. Both effects are visible within the latency range of the N400. Especially at lower frequencies (e.g., at theta frequencies) the temporal resolution of the wavelet transform is relatively poor. This implies that the relative onset difference between theta and gamma activity cannot be taken as a reliable indicator of onset differences in the underlying neurophysiological events. In particular the latency of the gamma peak seems to be later than the onset and the peak of the N400 ERP response. This suggests that the oscillatory brain responses reflect post-integration processes.

Gamma oscillations have been associated with feature binding within and across modalities (19, 20). These oscillations are suggested to play an important role in the integration of activity in both local and distributed cortical networks (17). It has been suggested that activity in the theta band might involve contributions from the hippocampal complex, with, presumably, additional cortical contributions (21). Theta activity has been observed in relation to both episodic and working memory tasks (22). Although the integration of word meaning and world knowledge occurs rapidly and simultaneously, the different oscillatory responses for semantic and world knowledge violations indicate that the brain seems to keep a record of the nature of the integration problem.

Finally, an event-related fMRI experiment was run using the same materials (23). The fMRI data, time-locked to the onset of the critical words, revealed an increase in activation in the left inferior prefrontal cortex (LIPC) as compared to correct sentences, that was common to both semantic and world knowledge violations (Fig. 3). The activation was observed in, or in the vicinity of, Brodmann’s areas 45 and 47. This region has previously been reported in connection to semantic processing (24, 25). This region is also known to contain one of the N400 generators, based on evidence from intracranial electrical recordings and MEG recordings (26, 27). Our study provides the first evidence for a role of the LIPC in the integration of world knowledge that is represented in long-term memory, next to the integration of lexical-semantic knowledge.

Both word meaning and world knowledge are recruited and integrated very rapidly, that is within some 400 ms, during on-line sentence comprehension. The left inferior prefrontal cortex seems to be critical both in the computation of meaning and in the verification of linguistic expressions. Although Frege (1892) made an important distinction between the sense of a proposition and its reference, their processing consequences appear to be immediate and parallel. Our results provide evidence against a non-overlapping two-step interpretation procedure in which first the meaning of a sentence is determined, and only then is its meaning verified in relation to our knowledge of the world. This is compatible
with findings of immediate influence of visual information (28), and of the preceding discourse on the interpretation of phrases and sentences (29). Semantic interpretation is not separate from its integration with nonlinguistic elements of meaning.

In conclusion, while reading a sentence, the brain retrieves and integrates word meanings and world knowledge at the same time. The left inferior prefrontal cortex is a crucial area for this integration process. Moreover, it does not take longer to discover that a sentence is untrue than to detect that it is semantically anomalous. However, the oscillatory brain responses suggest that the brain keeps a record of what makes a sentence hard to interpret, whether this is word meaning or world knowledge.

References and Notes
3. R. Jackendoff, Foundations of language: Brain, grammar, evolution (Oxford Univ. Press, Oxford, 2003). Although for Frege, reference was established relative to objects in the world, here we follow Jackendoff’s suggestion that this is done relative to objects and state of affairs as mentally represented.
10. Often word meanings can only be fully determined by invoking world knowledge. For instance, the meaning of “flat” in a “flat road” implies the absence of holes. However, in the expression “a flat tyre”, it indicates the presence of a hole. The meaning of “finish” in the sentence “Bill finished the book” implies that Bill completed reading the book. However, the sentence “The goat finished the book” can only be interpreted as the goat eating or destroying the book. The examples illustrate that word meaning is often undetermined and necessarily intertwined with general world knowledge. In such cases it is hard to see how the integration of lexical meaning and general world knowledge could be strictly separated. See (3) for a more extensive discussion; see for other examples: P. A. M. Seuren, Discourse semantics (Basil Blackwell, Oxford, 1985).
12. ERPs for 30 subjects were averaged time-locked to the onset of the critical words, with 40 items per condition. Sentences were presented word by word with a Stimulus Onset Asynchrony of 600 ms on the center of a computer screen. While subjects were reading the sentences, their EEG was recorded and amplified with a high-cut-off frequency of 70 Hz, a time-constant of 8 sec, and a sampling frequency of 200 Hz.
13. Information on EEG and fMRI methods and additional results are available on Science Online.
18. Time-Frequency Representations of the single trial EEG data were obtained by computing the squared norm of the result of the convolution of complex Morlet wavelets with the EEG data. We used wavelets with a 7-cycle width, with frequencies ranging from 1 to 70 Hz, in 1-Hz steps. Power values thus obtained were expressed as a percentage change relative to the power in a baseline interval, which was taken from 150 to 0 ms prior to the onset of the critical word. This was done in order to normalize for individual differences in EEG power, and for differences in baseline power between different frequency bands. Two relevant time-frequency components were identified: (i) a theta component, ranging from 4–7 Hz, and from 300–800 ms after word onset; (ii) a gamma component, ranging from 35–45 Hz and from 400–600 ms after word onset.
23. Whole brain T2*-weighted EPI-BOLD fMRI data were acquired with a Siemens Sonata 1.5 T MR-scanner (interleaved slice ordering, volume TR = 2.48 s, TE = 40 ms, 90° flip-angle, 31 horizontal slices, slice matrix = 64 × 64, isotropic voxel size = 3.5 × 3.5 × 3.5 mm³). For the structural MR image, a high resolution (isotropic voxels of 1 mm³) T1-weighted MP-RAGE pulse sequence was used. The fMRI data were preprocessed and analyzed using SPM99.
30. Supported by the Netherlands Organization for Scientific Research, grant 400-56-384 (PH). We thank P. Indefrey, P. Fries, P. A. M. Seuren, and M. van Turennout for helpful discussions.

**Supporting Online Material**

www.sciencemag.org/cgi/content/full/1095455/DC1

Materials and Methods

References and Notes

Fig. S1

8 January 2004; accepted 9 March 2004

Published online 18 March 2004; 10.1126/science.1095455

Include this information when citing this paper.

**Fig. 1.** Grand average (n = 30) ERPs for a representative electrode site (Cz) for correct condition (black line), world knowledge violation (blue dotted line), and semantic violation (red dashed line). ERPs are time locked to the presentation of the critical words (underlined). Semantic violations resulted in a larger N400 amplitude between 300 and 550 ms than the control condition \[F(1,29) = 73.0; P < 0.0001\]. N400 amplitudes to world knowledge violations were also larger than for correct controls \[F(1,29) = 27.4; P < 0.0001\]. The size of the effect was slightly larger for semantic violations than for world knowledge violations \[P < 0.05\]. The onset of the effects for semantic and world knowledge violations are not significantly different \([13]\). Spline-interpolated isovoltage maps display the topographic distributions of the mean differences from 300–550 ms between semantic violation and control (left); and between world knowledge violation and control (right). Topographic distributions of the N400 effect are not significantly different between semantic and world knowledge violation \(P = 0.9\).

**Fig. 2.** Morlet-wavelet based Time-Frequency Representations (TFR) of the induced activity for the three experimental conditions. The x axis represents time, the y axis frequency. Boxes within the TFRs for a representative electrode (Fz) indicate the areas in which clear gamma and theta responses were observed in a latency range overlapping with the N400. Left of the TFRs are the topographic maps of the gamma activity within the indicated area. On the right, scalp topographies are shown for the theta response in the indicated area. The world knowledge violation showed a peak in the gamma frequency range with a fronto-central distribution, with significant increase in gamma power for 8 electrodes (\(t\)-test, two-tailed, \(P < 0.05\)). No increases in gamma power were seen in the other conditions. All conditions showed a significant theta response. The theta response was significantly stronger for the semantic violation than for both other conditions \([F(1,29) = 4.2; P < 0.05]\), based on an analysis with left and right temporo-parietal, and frontal midline electrode sites as the regions of interest. In addition, only the semantic violation showed a frontal midline theta increase, as was found in a peak-surround analysis for midline electrode Fz \([F(1,29) = 11.2; P < 0.005]\).

**Fig. 3.** The common activation for semantic and world knowledge violations compared to the correct condition, based on the results of a minimum-T-field conjunction analysis. Both violations resulted in a single common activation \((P = 0.043, \text{corrected})\) in the left inferior frontal gyrus (in, or in the vicinity of Brodmann’s area 45 \([x, y, z] = [–44, 30, 8], Z = 4.87\) and BA 47 \([x, y, z] = [–48, 28, –12], Z = 4.15\). The cross-hair indicates the voxel of maximal activation and has the following coordinates \([x, y, z] = –44, 30, 8 \text{ (left BA 45)}\). L: left hemisphere; R: right hemisphere (supporting online text).
The Dutch trains are noisy and very crowded.

The Dutch trains are not noisy and very crowded.

Correct:

Semantic violation:

World knowledge violation: