PERSPECTIVES ON SENTENCE PROCESSING

edited by

Charles Clifton, Jr.
Lyn Frazier
Keith Rayner

PERSPECTIVES ON SENTENCE PROCESSING

This page intentionally left blank

PERSPECTIVES ON SENTENCE PROCESSING

Edited by
Charles Clifton, Jr.
Lyn Frazier
Keith Rayner
University of Massachusetts at Amherst

Psychology Press
Taylor & Francis Group
NEW YORK AND LONDON

First published 1994 by
Lawrence Erlbaum Associates, Inc.

This edition published 2016 by Psychology Press
711 Third Avenue, New York, NY, 10017, USA

and by Psychology Press
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Psychology Press is an imprint of the Taylor & Francis Group, an informa business

Copyright © 1994 by Lawrence Erlbaum Associates, Inc.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Library of Congress Cataloging-in-Publication Data

Perspectives on sentence processing / edited by Charles Clifton, Jr.,
Lyn Frazier, and Keith Rayner.
p. cm.
Selected papers from a conference held at the University of
Massachusetts at Amherst in March, 1993.
Includes bibliographical references and indexes.
1. Psycholinguistics—Congresses. I. Clifton, Charles, 1938–
II. Frazier, Lyn. 1952– . III. Rayner, Keith.
P37.P475 1994
401'.9—dc20 94-6096
CIP

ISBN 13: 978-0-805-81581-8 (hbk)
ISBN 13: 978-0-805-81582-5 (pbk)

Publisher's Note
The publisher has gone to great lengths to ensure the quality of this reprint but points out that some imperfections in the original may be apparent.

Contents

This page intentionally left blank

List of Contributors

Markus Bader Institut fuer Germanistische Sprachwissenschaft, Universitaet Jena, Leutragraben 1, UHH, 8, Stock, D-07743, Jena, Germany

Stephen Barton Department of Psychology, University of Glasgow, 56 Hillhead Street, Glasgow G12 9YR, Scotland.

Colin Brown Max Planck Institute for Psycholinguistics, P.O. Box 310, NL 6500 AH Nijmegen, The Netherlands.

Charles Clifton, Jr. Department of Psychology, University of Massachusetts, Amherst, MA 01003.

Cynthia M. Connine Psychology Department, State University at New York at Binghamton, Binghamton, NY 13760.

Laura Conway Department of Linguistics, University of Connecticut, Storrs, CT 06268.

Stephen Crain Department of Linguistics, University of Connecticut, Storrs, CT 06268.

Matthey W. Crocker Department of Artificial Intelligence, 80 South Bridge, Edinburgh EH1 1HN, Scotland.

Lyn Frazier Department of Linguistics, University of Massachusetts, Amherst, MA 01003.

Simon Garrod Human Communication Research Center, Department of Psychology, University of Glasgow, Glasgow G12 9YR, Scotland.

Edward Gibson Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139.

Peter Hagoort Max Planck Institute for Psycholinguistics, P.O. Box 310, NL 6500 AH Nijmegen, The Netherlands.

Aravind K. Joshi Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA 19103.

Ingeborg Lasser Max Planck Institute for Psycholinguistics, P.O. Box 310, NL 6500 AH Nijmegen, The Netherlands.

Simon P. Liversedge Department of Psychology, University of Dundee, Dundee DD1 4HN, Scotland.

Maryellen C. MacDonald Neuroscience Program, University of Southern California, Los Angeles, CA 90089–2520.

Wayne S. Murray Department of Psychology, University of Dundee, Dundee DD1 4HN, Scotland.

Weijia Ni Department of Linguistics, University of Connecticut, Storrs, CT 06268.

Lee Osterhout Department of Psychology NI-25, University of Washington, Seattle, WA 98195.

Neal J. Pearlmutter Beckman Institute, University of Illinois, 405 N. Mathews, Urbana, IL 61801.

Martin Pickering Department of Psychology, University of Glasgow, 56 Hillhead Street, Glasgow G12 9YR, Scotland.

Owen Rambow Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA 19103.

Keith Rayner Department of Psychology, University of Massachusetts, Amherst, MA 01003.

Julie Sedivy Department of Linguistics, University of Rochester, Rochester, NY 14627.

Edward P. Stabler Department of Linguistics, UCLA, 405 Hilgard Ave., Los Angeles, CA 90024.

Michael K. Tanenhaus Department of Psychology, University of Rochester, Rochester, NY 14627.

John C. Trueswell Department of Psychology, University of Pennsylvania, 3815 Walnut Street, Philadelphia, PA 19104

Pienie Zwitserlood Westf. Wilhelms-Universität Münster, Psychologisches Institut II, Fliednerstrasse 21, D48149 Münster, Germany

This page intentionally left blank

Preface

One of the liveliest forums for sharing psychological, linguistic, philosophical, and computer science perspectives on psycholinguistics has been the annual meeting of the City University of New York (CUNY) Sentence Processing Conference. The initial meeting of this conference was held at the CUNY Graduate Center in New York City in March 1988 and featured papers that represented nearly all the disciplines that contribute to psycholinguistics. Discussion of these papers repeatedly juxtaposed the insights of multiple disciplines, sometimes clashing, but more often complementing, one another. The next meeting of the conference added a poster session, at which participants learned how to distinguish the linguists' posters from the computer scientists' from the psychologists' by the style of their graphics and by the presence or absence of a pile of handouts. They also learned to dive into a poster from any discipline and expect to learn something interesting from it.

This interdisciplinary flavor has persisted through all the CUNY conferences that followed the initial one. The meeting of the sixth CUNY conference was held at the University of Massachusetts at Amherst in March 1993, giving the organizers at the CUNY Graduate Center a year's break. It held true to form. The papers and posters that were presented at this meeting represented the disciplines of linguistics, psychology, and computer science, many blending two or three disciplines in a single presentation. The organizers of this conference, who are the editors of this book, selected a subset of the presented papers and invited their authors to contribute them to a book. We intended to represent the main themes that ran through the 1993 conference, as well as honoring the breadth of presentations at the conference. We also hoped to highlight some of the most exciting current developments in the field of sentence processing, and even to glimpse what might be exciting in the next few years. The inability of some invited authors to provide a final manuscript by the tight deadline with which this book was prepared resulted in some narrowing of perspective and some underrepresentation of an important theme or two. Nonetheless, we hope readers agree that the chapters in this volume present the state of the art in several important approaches to sentence processing. We also hope the book rekindles in participants at the conference some of the excitement they felt when the conference was taking place.

The 1993 conference could not have taken place without the generous support of several institutions, including the University of Massachusetts/Amherst Training Program in Psycholinguistics and the University of Massachusetts/Amherst Graduate School, the Computational Linguistics Program of the Carnegie-Mellon University, the Institute for Research in Cognitive Science of the University of Pennsylvania, the Cognitive Science Program of the University of Arizona, and the Cognitive Science Program of the University of Rochester. Preparation of this volume was supported, in part, by grants to the University of Massachusetts (HD-07327 and HD-18708 from NIH, and DB-9121375 from NSF). The editors (both as editors and organizers of the conference) owe a great debt of thanks to Tom Maxfield, who helped organize and run the conference and pull this book together. We also thank Judi Amsel of Lawrence Erlbaum Associates for her help and encouragement in bringing this book to completion.

Charles Clifton, Jr.
Lyn Frazier
Keith Rayner

1 Introduction

Charles Clifton, Jr.
Lyn Frazier
Keith Rayner
University of Massachusetts at Amherst

The topics of how people produce and comprehend sentences and how they learn to do so have played a central role in the development of cognitive psychology and cognitive science. Demonstrations of the utter inadequacy of behavioral theory to deal with language comprehension, production, and acquisition (e.g., Chomsky, 1959; Miller, Galanter, & Pribram, 1960) led to its demise and fueled the cognitive revolution (Gardner, 1985; Johnson-Laird, 1988). A decade later, psycholinguistics was one of the core topics addressed by the new interdisciplinary field of cognitive science (Stillings et al., 1987). One reason that the psycholinguistics of sentence processing has played such a central role is the importance of language to the human state: How is it that we are so outstandingly successful at using language, given all our other inadequacies? Another reason is that the topic of sentence processing has proved to be a field that specialists in cognitive psychology, linguistics, computer science, and philosophy (and undoubtedly other disciplines) can study from their own perspective. In addition, it has proved to be a field in which researchers in one discipline can profit from the insights provided by researchers in the other disciplines.

This volume contains chapters that were presented at the sixth meeting of one of the liveliest forums for exploring interdisciplinary approaches to psycholinguistics—the City University of New York (CUNY) Sentence Processing Conference. We have organized the volume into six sections: (a) Sentence Processing and the Brain; (b) Phonological Processing; (c) Syntactic Processing: Information Flow and Decision Making; (d) Syntactic Processing and Computational Models; (e) Referential Processing; and (f) Sentence Processing and Language Acquisition. We provide a brief preview of the chapters in each of these sections, commenting just a bit here and there on how they fit into what is happening in the field as a whole. We conclude with some brief comments on current trends in sentence-processing research.

ORGANIZATION OF THIS BOOK

Sentence Processing and the Brain

Chapters on event-related potentials (ERPs) by Osterhout (chapter 2) and Hagoort and Brown (chapter 3) emphasize that ERPs provide an on-line technique not dependent on button pushing or any other secondary task. They also emphasize the multidimensional nature of the ERP waveforms and its continuous real-time nature. Both chapters show how syntactic violations result in components (P600 and sometimes left anterior negativity) that are distinct from the component known to attend semantic anomalies—the N400.

Osterhout notes that the P600 is apparent not only with phrase-structure and subcategorization violations, but also empty-category principle (ECP) and subjacency violations. The P600 begins earlier for these latter violation types, however, arguing against any temporal delay in the application of bounding constraints relative to phrase-structure constraints, and making the findings of Pickering, Barton, and Shillcock (chapter 9, this volume) concerning initial violation of island constraints quite surprising. Number- and gender-agreement violations also exhibit P600s, suggesting that they pattern with other syntactic violations. Both P600 and N400 are observed in examples where a syntactic violation results in an irrevocably ungrammatical and uninterpretable sentence. However, in garden-path sentences like "The boat floated down the river sank during the storm," the disambiguating word gives rise to either a P600 (for most subjects) or an N400 (for a few subjects), but never to both within a single individual.

The underlying source of the P600 versus N400 remains unclear, however, as Osterhout emphasizes. The P600 might directly reflect the processes underlying syntactic analysis, or the process of responding to syntactic violations, or processes of reanalysis. Hagoort and Brown echo this caution, and argue that certain violations (e.g., subcategorization) can be syntactic in origin but nevertheless have semantic consequences (e.g., on the interpretation of a lexical item's argument structure). They also present ERP data on lexical- and syntactic- ambiguity resolution, and they discuss how the ERP data compare with data obtained using other techniques.

Phonological Processing

Most work on sentence comprehension has used visual rather than auditory presentation. This can be justified by the intrinsic interest that the process of reading holds. Nonetheless, the real reason for the emphasis on reading is methodological and technical. It is technically easier to prepare well-controlled visual than auditory stimuli, it is hard to know how to describe an auditory signal properly without a good theory of prosody, and the methods used to probe the reading process are far more advanced than those available for studying auditory-sentence comprehension.

The roadblocks to studying listening are disappearing. Advances in micro-computer technology have brought most laboratories the capability of preparing and presenting auditory sentences, interesting and promising theories of prosody are appearing, and theories of auditory-word recognition have advanced far enough to permit serious development of theories of how people comprehend sentences they hear. This is fortunate because the field of auditory-sentence comprehension has the potential of providing important new insights into sentence processing. To take just two possible avenues of exploration, questions of how sentence context affects auditory-word recognition and how prosody affects sentence processing seem to be of unquestionable interest.

The two contributors to this volume who deal with auditory-language processing, Zwitserlood (chapter 4) and Connine (chapter 5), do not tackle questions of auditory-sentence comprehension head on, although both of these authors are known for contributions they have made to such topics. Rather, they both emphasize basic questions of how individual words are dealt with. Zwitserlood extends the work she has done on the topic of priming in auditory-word recognition to build a bridge to questions of spoken-word production (naming). She notes that, just as auditory-sentence processing has lagged as a field behind visual-sentence processing, the entire field of language production has lagged behind the field of language comprehension (and for rather similar reasons). She reviews the auditory phonological priming literature, looking at both facilitation and inhibition in word recognition, and argues that little evidence exists for true perceptual facilitation, whereas a rich variety of inhibitory effects have been demonstrated. She also reviews the phonological priming literature concerning language production, finding substantial evidence for facilitatory effects. Finally, she sketches a series of new experiments, looking at priming effects in both comprehension and production. She argues that the dissociation of effects in comprehension versus production suggests that the tasks use the same representations of linguistic form, but recruit different processing mechanisms. Partial overlap between the form of a prime and the form of a target item can facilitate phonological encoding in production via activation of lexical representations of words similar to but distinct from a target, but this same lexical activation, together with the deactivation of forms that minimally mismatch a prime, will have negative consequences for comprehension.

Connine addresses the question of why people are so good at recognizing auditory words when the problem seems to be so hard in the abstract. Why don't small distortions in the speech signal eliminate the intended word from the cohort of words compatible with the signal (cf. Marslen-Wilson & Tyler, 1980)? Utilizing both cross-modal priming and generalized phoneme-detection techniques, Connine shows that the degree of phonological similarity between an input and a target, the confusability of a target with other lexical entries, and the degree of redundancy in a target item all affect how much a given input activates a target lexical entry. Connine also introduces the concepts of *vertical* and *horizontal similarity.* Vertical similarity is (roughly) the degree of match between an acoustic–phonetic event and the corresponding lexical representation, and horizontal similarity is (again, roughly) the summed similarity between a lexical representation and all of its immediate phonological neighbors. There appear to be clear interactions between these two variables. For example, a high degree of vertical similarity between a nonword prime and a lexical target seems to result in more activation of a target item when horizontal similarity of the target to its neighbors is low than when it is high. Connine concludes that the process of auditory-word recognition can be a relatively gradual one, in which a preferred candidate emerges at a rate modulated by both vertical and horizontal similarity.

Syntactic Processing: Information Flow and Decision Making

Some of the most stimulating work in sentence processing is coming out of the laboratories at the University of Southern California (USC) and the University of Rochester. This work is guided by a theoretical perspective that emphasizes how various factors activate competing linguistic representations and how a constraint-satisfaction process chooses among them. It is notable for the variety of demonstrations it has provided of how sentence processing is sensitive to a wide range of extrasyntactic factors. The chapters by MacDonald, Pearlmutter, and Seidenberg (chapter 6) and Trueswell and Tanenhaus (chapter 7), together with other chapters elsewhere in this volume, provide an extended summary of the current state of this work.

MacDonald et al. argue a point that seems to be shared by most people working in the Rochester/USC tradition: Lexical-ambiguity resolution and syntactic-ambiguity resolution basically work the same way. They deny the widely held view (see, e.g., Rayner & Morris, 1991) that different processes hold in these two domains because lexical representations are retrieved, whereas syntactic representations are constructed. MacDonald et al. argue that a variety of variables, including frequency and semantic context, affect both types of ambiguity resolution in a similar fashion. They provide a striking analysis of published studies of the resolution of "horse-raced" sentences (the reduced-relative vs. past-tense ambiguity), suggesting that whether context overrides the preference for the simpler and generally more frequent past-tense reading depends on whether the verbs that were used occur more frequently in print in the past-participle or the past-tense form. Verbs that are frequently used as past participles are interpreted as such when the context supports it; verbs that are used less than 60% of the time as past participles are initially interpreted as past-tense verbs.

Trueswell and Tanenhaus provide their own arguments for an interactive (they use the term *constraint-based*) model of sentence comprehension. Like MacDonald et al., they emphasize the richness of the information provided by prestored lexical items, and focus on how the activation of a prestored item is determined by its likelihood given the input. This orientation encourages them to examine the frequency and plausibility of various structures and how such variables relate to comprehension difficulty. Also, like MacDonald et al., they emphasize how parsing difficulty should be a graded effect, not the more discrete effect that serial parsing models such as the garden-path model (Frazier, 1987) seem to suggest. Their research has led them to look at previously studied constructions, such as *that*-complements, in much greater detail than has been done previously, and to uncover some subtle and intriguing effects of argument-structure frequency, lexical preferences for specific linguistic forms (e.g., presence vs. absence of an overt *that* complementizer), and thematic-structure preferences.

The empirical effects discovered by researchers such as MacDonald et al. and Trueswell and Tanenhaus must clearly be accounted for by any adequate model of sentence processing. As these authors make clear, their own theorizing is still in a very preliminary and formative stage. The types of theories they are developing clearly can be sensitive to effects of frequency and preexisting biases they are demonstrating. Whether they can simultaneously be sensitive to the refined distinctions captured by the grammars of human languages may turn out to be a question of whether grammar can be reduced to prestored lexical structures that are activated by input.

Gibson and Pearlmutter (chapter 8) present one kind of research that must be done in developing theories of the sort envisioned by MacDonald et al. and Trueswell and Tanenhaus. They examine the relationship between the frequency of different resolutions of an ambiguity involving noun phrase (NP) modification and the ease of comprehending these different constructions. They suggest that frequency in a corpus should reflect the same complexity that can be measured in a comprehension-time experiment, and they provide evidence for this claim. In doing so, they are forced to filter their corpus sample in a variety of ways, graphically illustrating the nontrivial difficulty of doing proper frequency counts. They provide a useful discussion of the difficulty of interpreting the direction of causality in the relationship between frequency and comprehensibility. They note that, although one could claim that more frequent constructions are comprehended more easily because they are more frequent, one could equally well (given our current knowledge) posit that common underlying processes could underlie both comprehension and production, resulting in a common source of processing difficulty that appears as lower frequency of production and increased difficulty of comprehension.

Pickering, Barton, and Shillcock (chapter 9) examine a different kind of syntactic decision making and propose a different kind of theory. They examine the processing of sentences with "moved" constituents ("filler-gap" constructions), although here, as in previous work, they assume a direct relation between displaced constituents and their subcategorizers. Thus, they interpret the evidence without reference to the notion of *gap* or *trace*. They present evidence that effects of a potential gap are found inside syntactic islands, such as a relative clause, as well as outside them. They argue that this supports a parsing mechanism that first overgenerates possible structures and then filters structures that violate island constraints, although without traces, it remains unclear what is overgenerated and filtered. Specifically, Pickering et al. show that reading times in both self-paced and eye-movement studies are longer for the verb *painted* in (1a) and (1c) than in (1b) or (1d) (where the position where the true gap would appear is represented by a ___).

(1) a. I realized what the artist painted the larger mural with ___ today.
b. I realized that the artist painted the larger mural with ___ skill.
c. I realized what the artist who painted the larger mural ate ___ today.
d. I realized that the artist who painted the larger mural ate ___ cakes.

This longer reading time may reflect the process of assigning a "filler" as an argument of a potential (but erroneous) argument assigner, or it may reflect very quick-acting processes of rejecting such an assignment. Pickering et al. discuss some differences between their findings and previous findings, including the position at which a "false gap" effect occurs (cf. Stowe, 1986). They also make some interesting proposals for further research, focusing on how the possibility of interpreting a string of words as a full clause or a potential end of sentence might affect filler-gap processing (Goodluck, Finney, & Sedivy, 1991).

Bader and Lasser (chapter 10) address the nature of the grammar that the human sentence-processing mechanism (HSPM) exploits. They propose that it is a principle-based grammar, but deny that the parser must be a head-driven licensing parser of the kind sometimes associated with principle-based parsing. They report a study of German parsing that shows that structure is postulated before the head of a phrase is received. They show that an NP is taken as subject of a verb-final complement clause before the other elements of the clause, including its head, have been encountered. Further, the NP is not taken as the direct object of an immediately following verb, but instead is initially taken as the subject of a subsequent verb. This result argues against the view that a constituent is structured with reference to the first available licensor because the NP in Bader and Lasser's sentences could have been licensed as object of the following verb. Concretely, German readers read sentences ending in (2a), in which the NP is taken as object of the first (more deeply embedded verb, V1), more slowly than sentences ending in (2B), in which the NP is taken as the subject of V2.

(2) ... daß NP PP V1 V2

a. $[_{CP2}$daß $[_{CP1}$ NP PP V1] V2] NP is object of V1

b. $[_{CP2}$daß NP $[_{CP1}$ PP V1] V2] NP is subject of V2

Bader and Lasser's findings argue against a view of parsing as licensing. They also show that a certain amount of top–down postulation of phrase structure is required in the human sentence-processing mechanism. Specifically, the verb phrase (VP) of the complement clause CP2 must be postulated before its head V (or I) is encountered, so that the NP can be taken as its subject.

Syntactic Processing and Computational Models

The chapters in this section contain additional new proposals about the structure of the HSPM that are motivated, in part, by a desire to design and implement computational models of sentence parsing. These chapters continue the discussion of the nature of the grammar that the HSPM exploits, begun in the previous section by Bader and Lasser—specifically whether it is a principle-based grammar (Crocker, chapter 11) or a tree-adjoining grammar (TAG; Rambow & Joshi, chapter 12). They also continue to address questions of whether phrase structure is projected: (a) strictly bottom–up using only information from the head of the phrase, (b) using any licensing information from the first available licensor, or (c) in a top–down fashion using requirements imposed by functional items but not lexical items. Also, they introduce new proposals concerning the processing of displaced constituents, further examining a topic introduced by Pickering et al.

Crocker's proposal (that the requirements of functional items such as the complementizer *that* give rise to top–down postulation) offers one means by which Bader and Lasser's results could be handled in a parser exploiting a government-and-binding (GB) or principle-based grammar. A complementizer like *daß* is required to take a clausal complement, which could lead to postulation of the complement clause before the head of the complement clause is encountered. Crocker's suggestion that the requirements of only functional items, not lexical content items, give rise to top–down node postulation is interesting, but we suspect that it may not be consistent with facts (see Adams, Clifton, & Mitchell, 1993, in particular).

Concerning the parsing of displaced constituents, Crocker proposes that the parsing system postulates traces according to the active trace strategy (ATS), which allows traces to be postulated before all lexical items preceding the trace in the terminal string have been encountered. This account fits naturally with certain intuitions about processing, but should be checked experimentally. Further, Crocker's ATS proposal raises interesting additional questions about the parsing of traces versus other empty categories such as *pro* (see DeVincenzi, 1991, for a system that treats traces like other kinds of empty categories).

Rambow and Joshi also deal with the processing of displaced constituents, drawing a sharp distinction between German topicalization, which they argue is constrained by the grammar, and long-distance scrambling, which they argue is constrained by the parsing mechanism. Thus, although there are constraints on topicalization, there is no bound on either the number of elements that may scramble or the number of clauses over which an element may scramble. Using a tree-adjoining grammar (TAG), their system parses in a bottom–up fashion, and thus contrasts with Bader and Lasser's conclusions. The passing of subcategorization restrictions between one verb and a governing verb permits the governing verb to take an additional argument—the scrambled argument. This allows each argument to be "unwrapped" and stored with the appropriate predicate. Because constituents must be held in memory until they may be unwrapped (i.e., combined with their argument assigner), multiple scramblings will rapidly lead to greater processing complexity and diminished acceptability judgments. The complexity metric that Rambow and Joshi propose correctly predicts that Dutch cross-serial dependencies will be easier to process than their center-embedded German counterparts (Bach, Brown, & Marslen-Wilson, 1986). Within German, it predicts the ease of extraposition, as well as the difficulty of long-distance topicalization. The results of their study provide a powerful argument for TAGs, which are crucial to the explanations that Rambow and Joshi proffer.

In his chapter, Stabler (chapter 13) proposes the bounded connectivity hypothesis, which claims that the processing complexity of a structure increases quickly when more than one relation of any given type connects a (partial) constituent α (or any element of α) to any constituent external to α. Assuming that A′ extraction of an NP with a particular case forms a relevant relation, extraction of two NPs with the same case is predicted to be unacceptable—even in a language like Hindi, where extraction of two NPs with distinct cases is acceptable. Examples like (3), attributed to Mahajan, support the prediction.

(3) ??? kis-ko^{1} rām-ne kis-ko^{2} t_1 t_2 kahā- ki sar dard h∈
who-ACC Ram-erg who-ACC tell that head pain is
Who did Ram tell that who has a headache?

Similar support for the bounded connectivity hypothesis is found in 'morphological' causatives, assuming that they are derived from V-raising in the syntax. Each causative verb is raised to amalgamate with the higher verb. Stabler notes that a bound on the number of morphological causatives in a single construction is found widely in V-initial languages (Amharic, Arabic), as well as V-final ones like Quechua. Sentence (4), to be compared with (5), comes from Quechua:

(4) *Riku-chi-chi-chi-ni (chi = causative)

(5) Riku-chi-chi-ni
see-make-make-1Sg
I have shown it.

Presenting new data, Stabler argues against a linguistic account of the bound on morphological causatives. Semantic explanations also fail to account for the bound on repeated application of the causative, given the acceptability of periphastic causatives where no V-raising occurs (e.g., "The president made the general make the sergeant make the private kill the reporter").

In addition to offering a psycholinguistic explanation for data that have not been discussed previously in the psycholinguistic literature, Stabler also explains Dutch and German V raising by appeal to the bounded connectivity hypothesis. Unlike Rambow and Joshi's concern, which is to distinguish the relative increase in difficulty of the German versus Dutch pattern with successively more embeddings, Stabler explains why the jump from two verbs to three verbs creates noticeable difficulty in both languages. Stabler emphasizes that the bounded connectivity constraint is a sufficient, not a necessary, cause of psychological complexity, thus leaving room for other operations to also influence complexity. His approach is a novel one that opens up an entirely new set of questions about the subtrees implicated in parsing and the relations that can connect them, as well as expanding the empirical coverage of psycholinguistic complexity metrics. It also implies a view of 'constructed' memory that is both novel and intriguing.

Referential Processing

One issue that has received a great deal of attention in the last few years is the extent to which contextual and pragmatic information influences on-line parsing decisions. Four chapters on referential processing deal with this issue.

Garrod (chapter 14) is concerned with the general issue of anaphor resolution. He points out that much of the psycholinguistic literature has treated anaphor interpretation as an isolated process that is only rather loosely related to the processes involved in interpreting sentences and discourse. On the other hand, he notes that, in the linguistic literature, different anaphoric devices have been associated with a range of discourse functions that signal different ways in which a sentence should be resolved. Garrod's goal is to reconsider some of the processing assumptions about anaphor resolution in light of various linguistic analyses.

Garrod describes a referential hierarchy for identifying three dimensions by which anaphors vary. The first is contextual presupposition: the degree to which the interpretation of an expression is solely determined by the linguistic context. The second is referential function: the degree to which an expression is used to maintain reference to focused antecedents. The third is antecedent identifiability: the degree to which an expression uniquely specifies its antecedent reference. In his chapter, Garrod describes a number of interesting experiments relative to these distinctions.

The other three chapters in this section are all concerned with how contextual information influences the resolution of temporary syntactic ambiguities. Articles by Rayner, Carlson, and Frazier (1983), Crain and Steedman (1985), and

Altmann and Steedman (1988) set the stage for the debate that is currently raging. Rayner et al. presented evidence suggesting that pragmatic information does not influence the initial parsing of a syntactically ambiguous sentence. Rather, they argued that parsing decisions are made on structural grounds and that contextual and pragmatic information influences reanalysis processes. On the other hand, Crain and Steedman, as well as Altmann, argued that contextual information can override structural decisions.

All participants in the debate concerning the extent to which contextual and discourse factors influence parsing decisions acknowledge that contextual information has an effect. At issue is whether or not such information guides the selection of the first structural hypothesis.

Murray and Liversedge (chapter 15), Sedivy and Spivey-Knowlton (chapter 16), and Spivey-Knowlton and Tanenhaus (chapter 17) continue this debate. Murray and Liversedge report the results of a series of experiments that lead them to come down on the side that referential context does not affect on-line, immediate parsing decisions. On the other hand, Sedivy and Spivey-Knowlton and Spivey-Knowlton and Tanenhaus report evidence leading them to the other conclusion: Contextual and referential information interact with lexical information to determine initial parsing decisions.

It is encouraging that there is now so much interest in the general issue of the influence of discourse information on parsing decisions. Given the wide range of results that have been reported, it seems most appropriate at the moment to determine the situations in which context does and does not have an influence on parsing, rather than continue the debate of *when* context has its impact.

Sentence Processing and Language Acquisition

Crain, Ni, and Conway (chapter 18) provide an extremely stimulating discussion of the relation between adult processing and the language-acquisition device (LAD). They argue that language would be unlearnable if kids assigned "minimal-commitment" interpretations to sentences. No positive evidence could force a revision of a minimal-commitment interpretation. Rather, children must assign maximally disconfirmable interpretations (e.g., letting *only* associate with the largest possible contrast set). However, they argue that adult processing strategies lead to analyses that amount to minimal-commitment interpretations, suggesting a dissociation between language acquisition and adult sentence-processing principles. This proposal is fascinating. It raises issues about the distinctness of LAD and the sentence-parsing module, as well as many issues concerning the semantic preferences exhibited by the adult sentence-processing system. One might have thought that the assignment of focus would determine the interpretation of *only*, and that pragmatic relevance, not minimal disconfirmability, might influence the interpretation of the contrast set for *only*. Regardless of the outcome of future studies of these questions, the Crain et al. proposal moves the field ahead by being so bold and stimulating.

APPROACHES TO RESEARCH IN SENTENCE PROCESSING

Most readers of this volume are aware that there is now a wide variety of methodologies used to study sentence processing. This is very encouraging, and we hope that converging evidence emerges. In the past, the resolution of some critical theoretical issues has seemed to revolve around different methodologies being used. To take an obvious example, early research demonstrating that contextual and pragmatic factors do not influence initial parsing decisions utilized eye-movement data (see Ferreira & Clifton 1986; Rayner et al., 1983) and some self-paced reading data (Ferreira & Clifton, 1986), whereas evidence in favor of contextual information influencing such decisions was based exclusively on self-paced reading data (see Altmann & Steedman, 1988; Taraban & McClelland, 1988). Self-paced reading does not give the experimenter as much opportunity to discriminate first-pass effects from reanalysis effects. That is, the form of self-paced reading that is most diagnostic (see Rayner, Sereno, Morris, Schmauder, & Clifton, 1989) involves the presentation of one word at a time. Hence, readers cannot look back in text and must process each word more completely when first reading it. Therefore, arguments concerning the effects of contextual information on parsing had an inherent confound: Data against the efficacy of such information were based largely on one methodology, whereas data on the other side of the debate were based on a different methodology.

As the chapters in this volume make clear, there are now a number of laboratories that have eye-movement recording devices to study sentence processing. In addition, it is encouraging that ERP studies are now dealing with sentence processing (see the chapters by Osterhout and Hagoort and Brown). As we pointed out earlier, the roadblocks to studying sentence processing as people listen to sentences are disappearing. We think all of these developments are positive because they allow for the possibility of obtaining converging evidence to examine how people comprehend sentences. Thus, perhaps a clearer picture of various issues will emerge by examining results from eye-movement studies, self-paced reading studies, ERP studies, and listening experiments.

At a more general level, sentence-processing research has always involved interdisciplinary work: Psychologists, linguists, computer scientists, and philosophers (and others) have deep and abiding interests in language processing. Researchers from these different disciplines often bring differing theoretical orientations to the field, moving the study of language processing forward by focusing on different aspects of the problem.

Besides reflecting differences in theoretical orientation, the chapters in the present volume exemplify how researchers can have very different theoretical goals. An interesting dichotomy is developing between various researchers interested in language processing. Some think that theories should be molded by data, so that their theories are often at a formative stage. Others think that theories should be explicitly formulated and then tested against data. The chapters in this volume reflect both sides of this dichotomy. In the end, the differences in the way psycholinguists choose to investigate issues related to sentence processing may be primarily a matter of taste, but it is clear that the goal of the work described in this volume is to arrive at a comprehensive theory of language processing.

REFERENCES

Adams, B. C., Clifton, C., Jr., & Mitchell, D. C. (1993). *Lexical guidance in sentence processing: Further evidence for a filtering account.* Manuscript submitted for publication.

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition, 30,* 191–238.

Bach, E., Brown, C., & Marslen-Wilson, W. (1986). Crossed and nested dependencies in German and Dutch. *Language and Cognitive Processes, 1,* 249–262.

Chomsky, N. (1959). Review of *verbal behavior. Language, 35,* 26–58.

Crain, S. & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological parser. In D. Dowty, L. Kartunnen, & A. Zwicky (Eds.), *Natural language parsing* (pp. 320–358). Cambridge: Cambridge University Press.

DeVincenzi, M. (1991). *Syntactic parsing strategies in Italian.* Dordrecht: Kluwer Academic Publishers.

Ferreira, F., & Clifton, C., Jr. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25,* 348–368.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance* (pp. 559–586). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution.* New York: Basic Books.

Goodluck, H., Finney, M., & Sedivy, J. (1991). Sentence completeness and filler-gap dependency parsing. In P. Coopmans, B. Schouten, & W. Zonneveld (Eds.), *OTS Yearbook 1991* (pp. 19–31). Utrech, The Netherlands: University of Utrecht Press.

Johnson-Laird, P. N. (1988). *The computer and the mind.* Cambridge, MA: Harvard University Press.

Marslen-Wilson, W. D., & Tyler, L. (1980). The temporal structure of spoken language comprehension. *Cognition, 8,* 1–72.

Miller, G. A., Galanter, E., & Pribram, K. (1960). *Plans and the structure of behavior.* New York: Holt.

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior, 22,* 358–374.

Rayner, K., & Morris, R. (1991). Comprehension processes in reading ambiguous sentences: Reflections from eye movements. In G. B. Simpson (Ed.), *Understanding word and sentence* (pp. 175–198). Amsterdam, The Netherlands: Elsevier Science Publishers.

Rayner, K., Sereno, S., Morris, R., Schmauder, R., & Clifton, C., Jr. (1989). Eye movements and on-line language comprehension processes. *Language and Cognitive Processes, 4,* 21–50.

Stillings, N. A., Feinstein, M. H., Garfield, J. L., Rissland, E. L., Rosenbaum, D. A., Weisler, S. E., & Baker-Ward, L. (1987). *Cognitive science: An introduction.* Cambridge, MA: MIT Press.

Stowe, L. (1986). Parsing wh-constructions: Evidence for on-line gap location. *Language and Cognitive Processes, 1,* 227–246.

Taraban, R., & McClelland, J. R. (1988). Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectations. *Journal of Memory and Language, 27,* 597–632.

I SENTENCE PROCESSING AND THE BRAIN

This page intentionally left blank

2 Event-Related Brain Potentials as Tools for Comprehending Language Comprehension

Lee Osterhout
University of Washington

Even a cursory consideration of language comprehension is likely to lead to several conclusions about the psychological processes underlying comprehension. Perhaps most notably, comprehension is remarkably rapid, occurring essentially in "real time." Furthermore, despite its rapidity (and despite our intuitions to the contrary), comprehension is not instantaneous, but is instead a continuous process distributed over time. And if one assumes that formal descriptions of language provide even a rough approximation of the informational types and representations that are functionally involved in comprehension, then language comprehension must involve multiple levels of analysis (e.g., phonological, lexical, syntactic, semantic, pragmatic). The results of these multileveled analyses are then somehow integrated into a single coherent interpretation with incredible rapidity. Finally, these processes occur largely outside of our conscious awareness, and indeed remain (for the most part) inaccessible to consciousness.

A substantive model of language comprehension explains how all of this happens. But it is precisely because comprehension is a rapid, multileveled, unconscious process that gaining a substantive understanding of it has proved so difficult (cf. Swinney, 1981, 1982). As eloquently noted by Swinney (1981), to understand comprehension, we must examine the process as it occurs in real time, rather than describing the end-state results of comprehension (i.e., memory representations) or the underlying structure of the language (as provided by linguistic theory). One might surmise that the ideal tool for examining language comprehension mirrors the properties of comprehension itself. Such a tool would combine on-line, continuous, and nonintrusive measurement with a differential sensitivity to events occurring at distinct levels of analysis. Such a tool would also not rely on overt, conscious judgments made by the comprehender.

These ideal properties pose a formidable methodological challenge. Unfortunately, most of the commonly used methods lack many or all of these crucial properties. (One notable exception is the use of eye tracking to monitor reading; cf. Rayner, Sereno, Morris, Schmauder, & Clifton, 1989.) Many of the available methods involve measurements made *after* the process (e.g., word, phrase, or sentence reading time; grammatical judgments) or at a discrete moment *during* the process (e.g., the cross-modal priming technique; cf. Swinney, Onifer, Prather, & Hirshkowitz, 1979). These measurements typically require the use of a secondary task (e.g., button pressing) in addition to the primary task of language comprehension, and the secondary task often requires a conscious decision on the part of the subject (e.g., Is this stimulus a word?). Finally, none of the commonly used measures has been demonstrated to be differentially sensitive to events occurring at distinct levels of analysis. These measures typically respond similarly to events at any level. For example, sentence-reading times increase in the presence of either syntactic or semantic anomaly.

One method that at least in principle approximates the ideal tool as outlined is the recording of event-related potentials (ERPs) elicited during comprehension. ERPs are negative and positive voltage changes in the ongoing electroencephalogram that are time-locked to the onset of a sensory, motor, or cognitive event. Certain negative- and positive-going deflections in the ERP waveform (called *components*) have been shown to be sensitive to specific cognitive processes (for review, see Hillyard & Picton, 1987). The advantages of ERPs as measures of real-time cognition are clear: ERPs provide a continuous, on-line record of the brain's electrical activity that occurs during the process under study. Measurement of ERPs requires neither a potentially contaminating secondary task nor a conscious judgment on the part of the subject. And since ERPs provide at least a rough estimate of localization and lateralization of brain activity, they offer the added prospect of tying behavior and cognitive models more closely to brain function.

Of course, these salutary properties of ERPs are irrelevant unless it can be shown that ERPs are sensitive to the process of interest. The goal of this chapter is to selectively review evidence that ERPs are, in fact, quite sensitive to certain language-related events. Most intriguingly, some of this evidence allows one to speculate that ERPs (perhaps uniquely among current methods) are indeed differentially sensitive to events occurring at distinct levels of linguistic analysis during comprehension. This review focuses on recent findings, particularly those relating to syntactic aspects of comprehension. ERP studies of phonological processes (e.g., Kramer & Donchin, 1987; Polich, McCarthy, Wang, & Donchin, 1983; Rugg, 1985), repetition priming (Rugg, 1984a, 1985b), and semantic priming (e.g., Bentin, 1987; Bentin, McCarthy, & Wood, 1985) are not reviewed. For a general overview of the N400 component of the ERP, see one of several excellent recent reviews (Kutas & Van Petten, 1988; Fischler, 1990; Fischler & Raney, 1991).

METHODOLOGICAL CONSIDERATIONS

A few words concerning ERP methodology are in order. One methodological issue concerns the strategies available for studying cognition with ERPs. ERP researchers typically adopt one of two strategies (see also Coles, Gratton, & Fabiani, 1990; Kutas & Van Petten, 1988; Osterhout & Holcomb, 1993). The first strategy is to identify as precisely as possible the cognitive processes underlying some known ERP component. This can be accomplished (in principle) by determining the necessary and sufficient conditions for altering the component's waveform characteristics (amplitude and latency). The benefits of this strategy are substantial: With an electrophysiological marker of some cognitive process in hand, one can infer changes in the underlying cognitive process directly from changes in the ERP component. For example, prior work led Van Petten and Kutas (1987) to conclude that the amplitude of one ERP component, the N400, reflects a word's "activation level" in memory. More specifically, they concluded that highly activated words elicit small N400s, whereas less activated words elicit large N400s. These reasonable conclusions allowed them to investigate the effects of context on the processing of polysemous words, by measuring the N400s elicited by target words related to the contextually appropriate or inappropriate meanings of a polysemous word (e.g., "The gambler pulled an ace from the bottom of the DECK," followed by the target words *cards* or *ship*). The contextually inappropriate target words elicited a larger amplitude N400 than did the contextually appropriate target words, and (within the window normally associated with the N400) ERPs to inappropriate targets were indistinguishable from ERPs to control words that were not related to any meaning of the ambiguous word. Given the assumptions of Van Petten and Kutas concerning the processes underlying N400, these results seem to indicate that the contextually appropriate meanings of the polysemous word were selectively activated in memory. Although this approach to ERP research has considerable appeal, the mapping between changes in an ERP component and putative cognitive processes is often far from transparent. Importantly, experimental designs that assume knowledge of underlying cognitive processes carry with them the significant risk associated with a misidentification of these processes. If N400 amplitude reflects some aspect of cognition other than word activation, the set of interpretations one would consider in explaining the Van Petten and Kutas findings would probably change dramatically.

The second strategy for using ERPs to study cognition is to use a known ERP component to study some cognitive process even if the cognitive and neural events underlying that component are unknown. All that is needed for this approach to work is that some component must be shown to systematically covary with manipulations of stimuli, task, or instructions that are known or posited to influence the process under study. Having uncovered such a covariation, one can make certain inferences about relevant psychological processes based on between-condition differences in the ERP waveform. For example, several researchers have observed a slow positive-going wave (labeled the *P600* by Osterhout & Holcomb, 1992) in the ERP response to syntactically anomalous words. The specific cognitive events underlying the P600 are not known, and there is scant evidence that the P600 is in any sense a direct manifestation of syntactic processes. Regardless, all that is needed for the P600 to act as a useful tool for investigating syntactic analysis is evidence that the P600 reliably co-occurs with a syntactic anomaly. One can then use the P600 as an electrophysiological indicator of syntactic processing difficulty. The work reviewed next has, by and large, adopted this second approach to ERP research.

Another important methodological issue concerns the temporal information inherent in ERPs. Given the continuous, on-line quality of ERPs, they promise to reveal a great deal about the timing and ordering of language-related events. The critical temporal marker is often the moment in time at which the ERPs from two conditions begin to diverge significantly, rather than the peak latency of a particular component (see also Fischler, 1900; Osterhout & Holcomb, 1993). For example, the peak of one language-sensitive component, the N400, reliably occurs at about 400 ms after presentation of a word. Furthermore, the peak amplitude of the N400 is reliably larger for contextually inappropriate words than for contextually appropriate words (Kutas & Hillyard, 1980). However, divergences in the waveforms elicited by appropriate and inappropriate words typically emerge around 200–250 ms following word onset. The importance of this distinction becomes clear when considering whether N400 amplitude reflects processes associated with lexical access. If the peak amplitude of the N400 is taken as the relevant temporal marker, most researchers would argue that the component occurs too late to reflect lexical access (see, e.g., Sabol & De Rosa, 1976). However, if the onset of divergences in waveforms is taken to be the relevant temporal marker, the N400 is much closer to the temporal window thought to be associated with lexical access.

What sorts of inferences, then, can one make about the timing of cognitive processes based on ERP data? In practice, such inferences are fraught with danger, particularly if the ERP effects are relatively late occurring. For example, consider again the P600 effect associated with a syntactic anomaly. This effect often has an onset around 500 ms after presentation of the anomaly (Osterhout & Holcomb, 1992; Osterhout, Holcomb, & Swinney, in press; Osterhout & Mobley, 1993). This finding by no means licenses the inference that the assignment of syntactic roles to words occurs only 500 ms after word presentation. The P600 might reflect syntactic processes only indirectly; hence, the onset of the effect could be temporally distant from the syntactic processes themselves. Fortunately, very early onsets of ERP effects can sometimes license reasonably strong inferences about the timing of cognitive processes. A good example of this is provided by Holcomb and Neville (1991), who reported that the ERPs to contextually inappropriate words in spoken sentences begin to diverge from those to contextually appropriate words long before the entirety of the word has been encountered by the listener. This indicates quite clearly that the interaction between word recognition and context occurs long before the word can be recognized based on the acoustic input alone.

A separate issue concerns the use of items analyses as a procedure for generalizing the results beyond the particular set of items used in the experiment (cf. Clark, 1973). Although such analyses have become a standard procedure within psycholinguistics, they are rarely performed in ERP research. One reason for this is related to the signal-to-noise issue inherent in the signal-averaging procedure used to obtain the ERP (cf. Hillyard & Picton, 1987). In the subjects analyses, ERP researchers studying language-related components typically average over a minimum of 30 trials for each subject to extract the "signal" (the ERP) from the "noise" (randomly occurring EEG). This provides a sufficient signal-to-noise ratio when the ERP effect of interest is reasonably large (e.g., greater than 2 μV). To obtain an equivalent signal-to-noise ratio in the items analyses, experimenters would be required to run at least 30 subjects per experiment—a number far greater than that necessary to derive stable, reliable ERP averages over subjects. (In practice, the required number of subjects is likely to be greater than the number of items, for several reasons; e.g., the fact that between-subject variance is almost always greater than within-subject variance.) Running the required number of subjects for items analyses is often deemed prohibitively expensive in terms of the use of resources. Nor is the signal-to-noise issue the only problem associated with items analyses. For example, such analyses typically require tremendous quantities of computer memory and disk space—quantities that are simply too large for the typical ERP laboratory. Faced with these and other equally severe problems, most ERP researchers rely on replications across different sets of items to determine the generalizability of the effects of interest.

Finally, an important methodological concern in any examination of sentence processing is the mode of stimulus presentation. Because measurement of ERPs does not require an intrusive secondary task, in principle this method allows presentation to more closely approximate a "normal" comprehension environment. Indeed, several researchers have recently recorded ERPs during the comprehension of sentences presented as natural, continuous speech (Holcomb & Neville, 1990, 1991; Osterhout & Holcomb, 1993). Others have linked ERP measurements with eye saccades during "normal" reading (Marton & Szirtes, 1988). However, the standard method in most of the existing literature (and in most of the work reviewed here) has involved the visual presentation of sentences in a sequential, word-by-word manner, with typical word-onset asynchronies ranging from 300 to 650 ms. This stimulus presentation mode has been chosen because it allows an examination of an extended period of ERP activity to individual words that is uncontaminated by the ERPs to subsequent words. Although this precaution was reasonable as a first step, more recent work (e.g., Osterhout & Holcomb, 1993) indicates that it is not always necessary.

ERPS AND LANGUAGE

Correspondence Between Formal Theories and Comprehension Processes

One fundamental issue confronting psycholinguistic attempts to model language comprehension concerns the correspondence between formal theories of language and the cognitive-neural processes underlying comprehension. Linguistic theories of grammatical structure often distinguish among several levels of analysis (e.g., phonetic, syntactic, semantic, etc.). Perhaps the most basic distinction is that between syntax (sentence form) and semantics (sentence content). From a linguist's point of view, sentences that violate syntactic constraints (e.g., "John slept the bed") are quite distinct from sentences that violate semantic or pragmatic constraints (e.g., "John buttered his bread with socks"). Whether or not these levels of description apply to the psychological processes underlying language comprehension remains a point of dispute. A common assumption in much recent psycholinguistic work is that separable processes derive distinct syntactic and semantic representations of a sentence (cf. Berwick & Weinberg, 1983; Fodor, Bever, & Garrett, 1974). However, a popular alternative view (one that has gained considerable ground with the advent of "neural net" models) is that semantic interpretations can be derived directly, without an intervening syntactic level (Ades & Steedman, 1982; Bates, McNew, MacWhinney, Devescovi, & Smith, 1982; Johnson-Laird, 1977; McClelland & Kawamoto, 1986; Riesbeck & Schank, 1978).

This fundamental question has been difficult to address with standard measures, largely because, as noted, these measures tend to respond similarly to anomalies at different levels (Fischler & Bloom, 1980; Rayner et al., 1989; Stanovitch & West, 1983; Wright & Garrett, 1984). The multidimensional nature of ERPs might make them a more efficacious tool for addressing this issue, given two key assumptions. One assumption is that the processes associated with a given level of analysis are distinct from those associated with other levels. A second assumption is that cognitively distinct processes are mediated by neurally distinct brain systems. Given these assumptions, evidence that syntactic and semantic anomalies elicit dissimilar patterns of brain activity could be construed to support the claim that separable syntactic and semantic processes exist (Neville, Nicol, Barss, Forster, & Garrett, 1991; Osterhout & Holcomb, 1992).

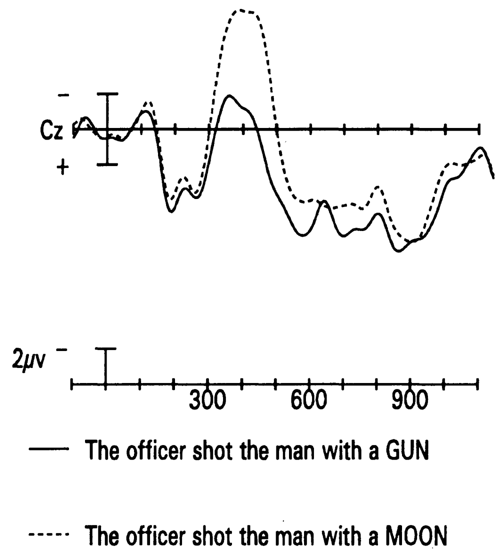

FIG. 2.1. Grand-average ERPs (averaged over subjects and items) recorded over site Cz to semantically anomalous words (dashed line) and nonanomalous control words (solid line) embedded within sentences. Onset of the critical words is indicated by the vertical calibration bar. Each hashmark on the horizontal axis represents 100 ms. Negative voltage is plotted up. Adapted from Osterhout (1990).

The pioneering work of Kutas and her associates over the past 15 years has demonstrated that the brain's electrophysiological response is measurably sensitive to at least one form of semantic-pragmatic anomaly. Kutas and Hillyard (1980a, 1980b, 1980c, 1983, 1984) found that contextually inappropriate words elicit a large-amplitude negative-going wave with a peak amplitude around 400 ms poststimulus (the N400 component; see Kutas & Van Petten, 1988, for review). A typical N400 response, elicited by words like *moon* in the sentence "The officer shot the man with a *moon* last night," is shown in Fig. 2.1. Although the precise cognitive (and neural) events underlying the N400 remain unclear, N400 amplitude appears to be a function of the semantic fit between the target word and preceding context.

Subsequent work has left little doubt that N400 amplitude is (at least in many situations) associated with semantic aspects of comprehension. The question, then, is whether the N400 also functions as a metric of syntactic processes, or if such processes are associated with a different (or any) ERP component. This question was addressed by Osterhout and Holcomb (1992). We presented sentences containing apparent violations of verb subcategorization or violations of phrase-structure constraints, as in (1) and (2):

(1) The broker hoped *to* sell the stock.
(2) *The broker persuaded *to* sell the stock.

The clausal complement "to sell the stock" can be easily attached to the fragment "The broker hoped" in Sentence (1). However, when used in its active form, the

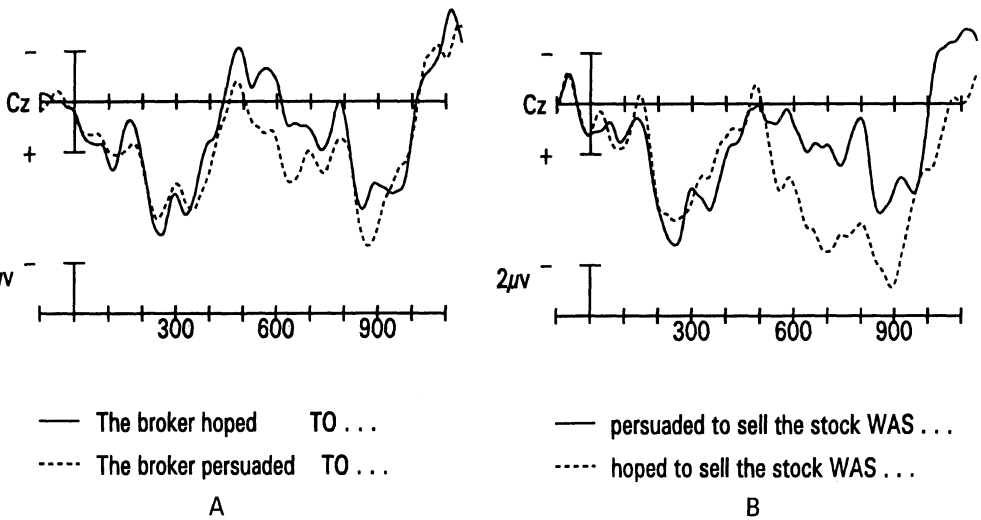

FIG. 2.2. (A) ERPs to apparent violations of verb subcategorization (dashed line) and grammatical control words (solid line). (B) ERPs to phrase-structure violations (dashed line) and grammatical controls words (solid line). Adapted from Osterhout and Holcomb (1992).

transitive verb *persuade* does not allow an argument beginning with the word *to* (i.e., a prepositional phrase [PP] or an infinitival clause) to occur immediately adjacent to the verb. Hence, the infinitival marker *to* in (2) is likely to be perceived as a violation of verb subcategorization, assuming readers do not initially attempt a reduced-relative clause analysis of the sentence. Fig. 2.2A plots the ERPs elicited by the infinitival markers in (1) and (2). The waveforms elicited by the grammatical and ungrammatical sentences clearly diverge between roughly 500 and 800 ms following presentation of the infinitival markers. However, the response to the syntactically anomalous case is not the negative-going N400, but is instead a positive-going wave. Because the positivity had a midpoint around 600 ms, we labeled this wave *P600*.

This finding suggests that a syntactic anomaly does elicit a brain response, but that this response is quite distinct from the N400 elicited by a semantic anomaly. In a second study, we presented lengthened versions of the sentences used in our first experiment, as illustrated in (3) and (4):

(3) *The broker hoped to sell the stock *was* sent to jail.

(4) The broker persuaded to sell the stock *was* sent to jail.

In (3) the added phrase cannot be attached to the initial part of the sentence without violating the phrase structure rules of English. Thus, (3) is ungrammatical and becomes so at the auxiliary verb *was*. Conversely, the added phrase can be attached to the initial part of Sentence (4); the verb *persuade* can be passivized, and this allows a reduced-relative clause interpretation of the sentence (corresponding to "The broker [who was] persuaded to sell the stock was sent to jail"). Under such an analysis, the auxiliary verb can be attached as part of the main clause ("The broker was sent to jail") and a syntactic anomaly is avoided. Therefore, if the P600 is associated with a syntactic anomaly, then the word *was* in (3) should elicit a P600 relative to the ERPs to the same word in sentences like (4). ERPs to the auxiliary verbs in both sentences types are shown in Fig. 2.2B. As predicted, the auxiliary verbs in the ungrammatical sentences elicited a large P600. These words also elicited a left-hemisphere, anterior negativity between 300 and 500 ms (not observable in Fig. 2.2B). Similar results have been reported by Hagoort, Brown, and Groothusen (1993), who used Dutch sentences as stimuli.

The basic distinction between syntax and semantics is not the only relevant distinction made by formal theories of grammar. For example, government-and-binding (GB) theory (Chomsky, 1981, 1986) posits the existence of multiple modules of grammatical knowledge within the syntactic domain. GB includes one module for specifying constraints on the phrase structure of sentences (X-bar theory) and other modules (theta theory, case theory, subjacency, empty category principle, etc.) that constrain "movement" of sentence elements for question and relative-clause formation. It is conceivable that a direct mapping exists between the grammar and the comprehension system, such that these modules of grammatical knowledge (or some subset of these modules) are encoded in cognitively and neurally distinct processing systems. (For claims in this direction, see Ferreira & Clifton, 1986; Frazier, 1990; Freedman & Forster, 1985). If these modules of knowledge are, in fact, instantiated in distinct processing modules within the brain, then one prediction is that each type of violation will elicit a distinct brain response. Two recent studies have examined this possibility. McKinnon and Osterhout (1993) examined the response to violations subsumed under the subjacency principle and the empty-category principle (ECP). Examples of violations of these constraints are presented in Sentences (5) and (6):

(5) *I wonder which of his staff members$_i$ the candidate was annoyed *when* his son was questioned by ____ $_i$.
(subjacency violation)

(6) *John$_i$ seems *that* it is likely ____ $_i$ to win.
(ECP violation)

The underlined word in each sentence marks the first point at which subjects might note the ungrammaticality of these sentences. The blank spaces represent "gaps" formed by movement of a sentence constituent (the "filler") from its canonical position. The subscripts index the gap to the filler. It is sufficient for present purposes to note that, within the GB framework, (5) is ungrammatical

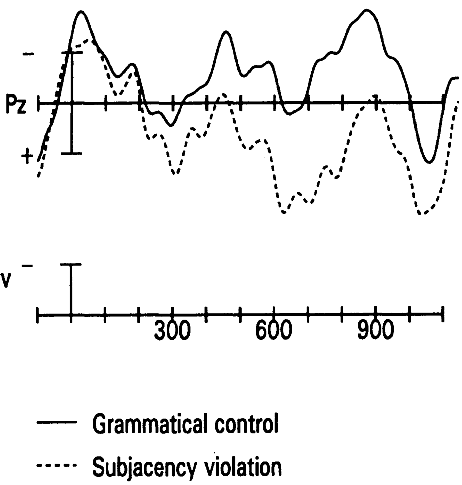

FIG. 2.3. ERPs to subjacency violations. (dashed line) and grammatical controls (solid line).

because the filler is too distant from its gap. More accurately, the structure intervening between the filler and the gap prevents a coindexation of the two (cf. Chomsky, 1981, 1986). Sentence (6) is ungrammatical because the gap does not maintain the proper structural relation to its filler or other proper governor (Lasnik & Saito, 1992).

The response to both types of anomalies was highly similar. (The response to subjacency violations is shown in Fig. 2.3.) Both anomalies elicited a positive-going wave similar in its component characteristics to the P600 effect reported by Osterhout and Holcomb (1992). The major difference was that the response to the subjacency and ECP violations had an onset around 200 ms. In contrast, the onset of the response to phrase-structure and subcategorization violations reported by Osterhout and Holcomb was not evident until about 500 ms.

Similar data have been reported by Neville et al. (1991). These investigators measured the response to violations of three distinct constraints: (a) phrase-structure rules, (b) subjacency, and (c) a third constraint known as the specificity constraint, which stipulates that a *wh*-phrase cannot be moved out of a noun phrase (NP) with specific reference (consider the unacceptability of "What$_i$ did the man admire Don's sketch *of* ____ $_i$?"). As in the Osterhout and Holcomb study, phrase-structure violations elicited both an enhanced negativity between 300 and 500 ms over left-hemisphere sites and a large positive-going wave beginning at about 500 ms. Subjacency violations elicited a positivity with an onset around 200 ms. Specificity violations elicited a slow negative-going potential that was most evident in the left hemisphere and that had an onset as early as 125 ms at some sites.

Whether or not these results indicate the operation of distinct processing modules has been a matter for debate. To date, all syntactic anomalies tested so far (excepting specificity violations) have elicited a broad-based positive-going wave (the P600), beginning between 200 and 500 ms after presentation of the anomaly. Assuming that similar brain responses reflect similar cognitive states, these data seem to indicate that many syntactic anomalies engender a similar processing response. This response is quite distinct from that elicited by a semantic anomaly (cf. McKinnon & Osterhout, 1993). At the same time, differences do exist among the responses to these different violations, particularly in the portion of the waveforms preceding P600 onset. For example, phrase-structure violations were associated with a left-hemisphere negativity not seen in the response to other anomalies, and P600 onset was much more rapid for subjacency violations than for phrase-structure violations. These differences can be construed as evidence that separable processing systems exist for each type of constraint (cf. Neville et al., 1991). Of course, there is nothing inconsistent in maintaining both claims simultaneously. It is certainly possible that, although the syntactic anomalies produce a common end state (reflected in the P600), this end state is reached via nonidentical sets of processes (reflected in the ERP differences preceding the P600).

The previous data exemplify situations in which formal theories of language structure have informed processing theories. One might ask whether opportunities exist for the converse situation to apply, that is, data from processing studies informing formal theory construction. In particular, one might conceivably use ERPs to provide an empirical, evidential basis for identifying the level of analysis at which certain phenomena occur within the processing system. As noted by Radford (1988), this is not always obvious. In examining the status of the anomalous sentences such as "The boy next door never loses her temper with anyone," Radford notes that the proper characterization of such oddities is open to argument. One could claim that the sentence is *syntactically* anomalous, that is, that agreement between a pronoun and its antecedent in gender is stipulated as part of the formal, rule-governed grammar. Alternatively, one could claim that the sentence is *semantically* anomalous, i.e., that part of the meaning of the noun *boy* denotes a male human, whereas part of the meaning of the pronoun *she* denotes a female human, leading to a contradiction in meaning if these two entities are taken to be co-referential. This ambiguity is mirrored in linguists' theoretical treatments of agreement (cf. Barlow & Ferguson, 1988). Traditional grammars (and many modern grammars) treat agreement as part of the syntactic (form-driven) system (Chomsky, 1986; Quirk & Greenbaum, 1973). But other recent accounts propose a semantic or discourse-function (content-driven) account of agreement (Givon, 1976; Reid, 1991).

Clearly, if our previous interpretations are correct, ERPs might prove helpful in deciding this issue from an empirical basis. The question is simple: Do violations of agreement elicit a brain response that is more similar to that elicited by a syntactic or semantic anomaly? That is, will such anomalies elicit the P600 (syntactic anomaly) or the N400 (semantic anomaly)? We (Osterhout & Mobley, 1993) recently addressed this issue by presenting sentences similar to (7)–(9):

(7) *Many doctors *claims* that insurance rates are too high.
(8) *The hungry guests helped *himself* to the delicious meal.
(9) *The successful woman congratulated *himself* on the promotion.

All three sentences contain an agreement violation. In Sentence (7), the verb disagrees with its subject in number. In (8) and (9), the reflexive disagrees with its antecedent in number and gender, respectively. Of interest are the ERPs elicited by the italicized words in each sentence. All three anomaly types elicited a large positive-going wave similar to the P600 component discussed previously. One could interpret these data as indicating that agreement is encoded as part of the syntactic, rule-governed constraints on well formedness. Similar results have been obtained with agreement violations that were presented to Dutch speakers (Hagoort et al., in press).

ERPs and On-Line Syntactic Analysis

One of the more notable trends within psycholinguistics over the past decade has been the resurgence of interest in the on-line syntactic analysis of sentences. This resurgence has been fueled largely by the seminal work of Frazier and Rayner (1982; Rayner, Carlson, & Frazier, 1983) and their associates (e.g., Ferreira & Clifton, 1986; Ferreira & Henderson, 1990). These researchers have compellingly demonstrated that, given an appropriate tool (eye tracking in this case), one can observe even those comprehension processes that occur with great speed. Because ERPs appear to be sensitive to syntactic aspects of comprehension, it becomes feasible to attempt to use ERPs as another on-line tool for investigating the psychological and neural processes underlying syntactic analysis during sentence comprehension.

One particularly productive line of research has entailed examining the processing response to syntactic ambiguity (i.e., situations in which more than one well-formed syntactic analysis is available for a string of words). For example, consider Sentence (10):

(10) The lawyer charged (that) the defendant *was* lying.

Without the complementizer *that,* the proper grammatical role of the NP "the defendant" is temporarily ambiguous between an "object of the verb" role and a "subject of an upcoming clause" role. The fact that the NP is actually the subject of a clausal complement becomes certain only when the syntactically disambiguating auxiliary verb *was* is encountered. In contrast, the presence of the overt complementizer immediately indicates (prior to encountering the NP) that the subject role is the appropriate role for the upcoming NP. Considerable evidence (mostly involving measurements of eye movements during reading) has indicated that readers experience processing difficulty upon encountering the auxiliary verb in sentences like (10) when no complementizer is present (Rayner & Frazier, 1987). These results have been taken to indicate that readers initially (and erroneously) assign the object role to the ambiguous NP (i.e., readers seem to pursue a single syntactic analysis even when confronted with syntactic ambiguity). Under a direct-object analysis, the phrase "was lying" cannot be attached to the preceding sentence fragment, and therefore a syntactic garden path results; the reader must attempt a reanalysis of the sentence.

Data from the few ERP studies that have examined this issue are, by and large, consistent with this serial parsing model. The results reported by Osterhout and Holcomb (1992), discussed earlier, are consistent with the claim that readers initially computed a simple-active analysis when confronted with a simple-active/reduced-relative clause ambiguity. Further work by Osterhout et al. (in press) provided additional evidence. Osterhout et al. presented sentences similar to Sentence (10). The auxiliary verb in sentences without an overt complementizer elicited a P600-like positivity, relative to ERPs to the same words in sentences with a complementizer. One could interpret this result as indicating that readers initially pursued the direct-object analysis of the reduced sentences, leading to a syntactic anomaly (a garden-path effect) upon encountering the disambiguating auxiliary verb.

If readers initially pursue a single analysis for syntactically ambiguous sentences, what factors determine which analysis is attempted first? Two distinct (and mutually exclusive) generalizations have been proposed regarding this question. Frazier and her colleagues have persuasively argued for a "minimal-attachment" strategy, in which the simplest analysis (as determined by the number of nodes in the phrase structure) is always attempted first, with backtracking and reanalysis when the minimal-attachment analysis turns out to be incorrect (Frazier, 1987; Frazier & Rayner, 1982). Other theorists have argued for a "lexical-preference" parser, in which the parser initially pursues the analysis that is consistent with the "preferred" subcategorization frame of the matrix verb in the sentence (Fodor, 1978; Ford, Bresnan, & Kaplan, 1982; Tanenhaus & Carlson, 1989). Implications of these two approaches are illustrated by Sentences (11)–(14):

(11) The doctor hoped the patient *was* lying.
(12) *The doctor forced the patient *was* lying.
(13) The doctor believed the patient *was* lying.
(14) The doctor charged the patient *was* lying.

These sentences are distinguished by the subcategorization properties associated with the main verb in each sentence. The intransitive verb *hope* in Sentence (11) does not allow a direct-object NP, unambiguously indicating that the NP is the subject of an upcoming clause. The transitive verb *force* in Sentence (12) requires a direct object, forcing the postverbal NP to play the object role. This results in ungrammaticality when the auxiliary verb *was* is encountered. The verbs in Sentences (13) and (14) can be used with or without a direct object. This introduces temporary syntactic ambiguity—the postverbal NP might be acting either as object of the verb or as subject of an upcoming clause. Because the object interpretation is syntactically simpler (Frazier & Rayner, 1982), a minimal-attachment parser would initially (and erroneously) assign the object role to the NP. This would lead to the apparent ungrammaticality due to a garden-path effect at the auxiliary in both sentences. In contrast, a lexical-preference parser would initially choose the analysis consistent with the verb's biases. The verb in Sentence (13) is biased toward intransitive use, whereas the verb in Sentence (14) is biased toward transitive use. Hence, a lexical-preference parser would initially pursue the correct verb–subject analysis of Sentence (13), but would erroneously pursue the verb–NP analysis of (14), leading to a garden-path effect at the auxiliary verb.

Sentences similar to Sentences (11)–(14) were visually presented (in a word-by-word manner) by Osterhout et al. (in press). ERPs to the final three words in each sentence type (postverbal noun, auxiliary verb, and the sentence-ending verb) are shown in Fig. 2.4. Arrows indicate the onset of each word. As expected, the auxiliary verbs in sentences containing "pure" transitive verbs elicited a large P600 effect, relative to the same words in sentences containing "pure" intransitive verbs. Less expectedly, these auxiliary verbs also elicited an enhanced N400 component. This observation is a challenge to our claim that the P600 and N400 effects are elicited as a function of syntactic and semantic anomalies, respectively. One reasonable explanation that allows us to maintain this claim hinges on the observation that the auxiliary verbs in these sentences rendered the sentence irrevocably ungrammatical (hence, uninterpretable). In contrast, the auxiliary verb in sentences like Sentences (10) and (14) simply force the parser to consider a less preferred analysis. Hence, ERPs to the auxiliary verb in Sentence (12) might contain the response to both a syntactic anomaly and the response to the semantic anomaly engendered by the ungrammaticality (and resulting uninterpretability) of the sentence.[1]

---

[1]Given such an interpretation, the relative onsets of the N400 (around 200 ms) and P600 (around 500 ms) become somewhat paradoxical. According to most accounts, syntactic analysis precedes semantic analysis (and indeed the semantic anomaly here is the result of the syntactic anomaly). Yet the brain response to semantic anomaly is actually preceding the response to syntactic anomaly. Since we cannot at present precisely identify the cognitive events underlying these effects, the resolution to the paradox is not clear. However, we would note that these effects may not directly reflect the processes involved in constructing syntactic and semantic representations; rather, it is likely that these

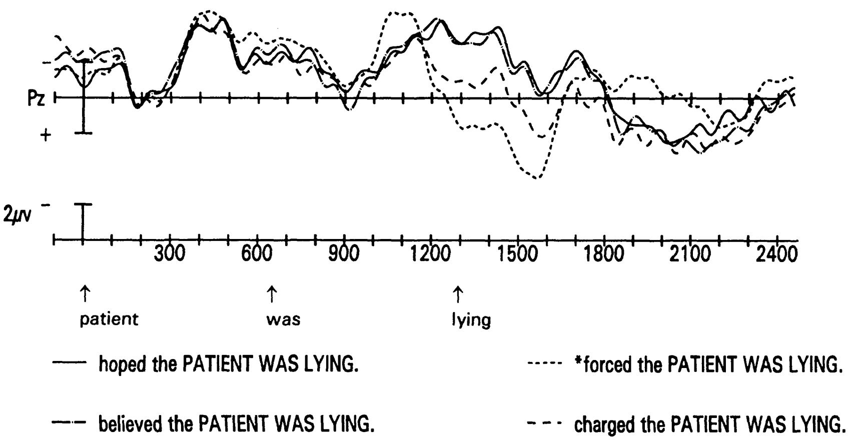

FIG. 2.4. ERPs to the final three words in each of the four sentence types: intransitive, transitive, intransitively biased, and transitively biased. Onset of the critical word (the auxiliary verb, penultimate in the sentence) is indicated by the second arrow under the horizontal axis. Adapted from Osterhout, Holcomb, and Swinney (in press).

Fig. 2.4 also indicates that auxiliary verbs in sentences containing transitively biased verbs also elicited a P600, although reduced in amplitude and more restricted in distribution than the P600 elicited by pure transitive-verb sentences. In striking contrast, ERPs to auxiliary verbs in sentences containing *intransitively* biased verbs were indistinguishable from those elicited by the same words in "pure" intransitive sentences (i.e., the auxiliary verbs in these sentences did not elicit a P600 effect). Clearly, P600 amplitude was a function of the verb-subcategorization information associated with the matrix verbs in these sentences. These findings suggest that the putative verb biases exist and have processing relevance at least under some comprehension environments.[2] Perhaps more importantly, these results provide compelling evidence that P600 amplitude is a function of syntactic aspects of comprehension.

Another type of syntactic ambiguity concerns the processing of "filler-gap" relations within sentences. As noted, these sentences contain a constituent (the filler) that has been "moved" from its canonical position within the sentence (the

components are indeterminately removed from the syntactic and semantic processes themselves. Also, this explanation is seemingly inconsistent with the results observed by Osterhout and Holcomb (1992). In that study, the phrase structure violations rendered the sentence irrevocably ungrammatical; yet, the response to the anomaly was a monophasic positivity.

[2]For a discussion concerning whether this evidence indicates that verb information is used to determine the *initial* syntactic analysis pursued by the parser or, alternatively, to aid in the re-analysis of an initial parse (as, e.g., predicted by models of parsing operating under a minimal attachment principle; cf. Frazier, 1987), see Osterhout et al. (in press).

gap). Considerable evidence indicates that the comprehension system attempts to match up fillers with the appropriate gaps during comprehension (cf. Bever & McElree, 1988; MacDonald, 1989; Nicol & Swinney, 1989; Osterhout & Swinney, 1993). However, this process of gap filling is far from trivial, as there is often uncertainty concerning the proper location of the gap. For example, the sentence fragment "The mother found out which book the child read. . . ." can be continued in several ways, each with a different gap location. Two models have been proposed to account for how the processing system deals with such uncertainty (cf. Fodor, 1978; Garnsey, Tanenhaus, & Chapman, 1989). A *first-resort parser* assigns a filler to the first possible gap, whereas a *last-resort parser* waits until there is unambiguous information about gap location. Garnsey et al. conducted a clever experiment to contrast these parsing models. They presented sentences similar to Sentences (15) and (16):

(15) The businessman knew which customer$_i$ the secretary called _____ $_i$ at home.
(16) The businessman knew which article$_i$ the secretary called _____ $_i$ at home.

The first possible gap location in these sentences is immediately after the verb (in direct-object position). However, the sentences could continue in such a way that the gap actually occurs at a different location (e.g., "The businessman knew which article the secretary called about _____"). A first-resort parser would posit a gap immediately after the verb, whereas a last-resort parser would wait until the proper gap location could be identified with certainty. The logic of the experiment was as follows: The noun *customer* is a plausible object of the verb *call,* whereas the noun *article* is not. If the parser immediately posits a gap after the verb and associates the gap with the filler, and if that filler is an implausible object, the verb might be expected to elicit an N400 component. This is precisely what Garnsey et al. observed, providing support for the first-resort parsing model.

Also relating to filler-gap sentences is the posited distinction between processes that derive phrase-structure representations (constituent-analysis processes) and those that determine relations among phrasal types (such as the "binding" between a gap and its antecedent). There are theoretical and empirical reasons for believing that these two tasks are distinct and separable (Frazier, 1990; Freedman & Forster, 1985; Forster, 1987). One theoretical reason for anticipating such a distinction within the processing system is that this distinction is explicit within GB theory. Furthermore, within GB there is an implied sequentiality in the operation of these processes; the phrase-structure representation must be constructed before relationships among phrasal constituents can be checked. This has led to the hypothesis that listeners and readers might "overgenerate" constituent structures that are locally well formed but that violate constraints on relationships among constituents. And indeed there have been claims of evidence purporting to show that readers "overgenerate" sentence structures in just this way (Freedman & Forster, 1985; Forster, 1987; Forster & Stevenson, 1987; but see Crain & Fodor, 1987). There is also evidence that certain aphasics can generate complete constituent structures, but cannot perform binding operations and other semantic processes over these structures (Linebarger, 1989).

If constituent analysis and binding operations are indeed separable and sequential, one might anticipate evidence that binding constraints are applied only after some measurable delay (cf. Weinberg, 1987). Some evidence bearing on this prediction is provided by the studies by McKinnon and Osterhout (1993) and Neville et al. (1991). In both experiments, sentences were presented that contained violations of constraints on movement (i.e., subjacency and ECP). As noted earlier, both types of anomaly elicited a P600-like positivity. Critically, the onset of this positivity was actually much more rapid for subjacency and ECP violations than for phrase-structure and verb subcategorization violations (compare Fig. 2.2B and 2.3).[3] These results are difficult to reconcile with the claim that binding constraints are applied after phrase-structure constraints.

Are These Language-Related ERP Effects Language Specific?

One of the standard doctrines within modern neuropsychology is the existence of language-specific brain systems (cf. Geschwind, 1979). Therefore, it is reasonable to ask whether the ERP components that are sensitive to language processes (e.g., N400 and P600) are in any sense language specific. Such claims have been made recently with respect to the N400 component (Holcomb, 1988; Holcomb & Neville, 1990). With respect to the P600, the most salient alternative to the language-specificity hypothesis is the possibility that the P600 is a member of the family of late positive components (P300 and related components) often observed following unexpected stimuli (Donchin, 1979, 1981; Duncan-Johnson & Donchin, 1977; Hillyard & Picton, 1987; Ritter & Vaughan, 1969). The amplitude of the P300 is a function of both the subjective probability and the task relevance of the eliciting stimulus. Because the P300 is elicited by a wide variety of stimuli, it is clearly not language specific. One could easily justify the claim that the syntactic anomalies in the experiments reviewed previously acted as "unexpected events," either by virtue of the general rarity of ungrammaticality or because readers generate expectations concerning upcoming sentence constituents.

[3]This comparison is somewhat complicated by the fact that in the Osterhout and Holcomb (1992) study, words were presented at a rate of 650 ms per word, while in the McKinnon and Osterhout study words were presented at a rate of 400 ms per word.

There are at least three ways to evaluate the possibility that the P600 is "just another" P300.[4] First, one can directly compare the scalp distributions of the P300 and P600 components. It is generally agreed that ERP components with distinct scalp distributions must be generated by neurally distinct brain systems (cf. Johnson, 1993). Second, one can attempt to determine if P600 amplitude is affected by the same manipulations known to affect P300 amplitude. If not, one has grounds for arguing that the effects are functionally distinct. Third, and most critically, one can determine whether the P600 and P300 components are *additive* in their effects. This can be accomplished by presenting stimuli that are expected to simultaneously elicit both components, and comparing this response to the ERP response to each anomaly type in isolation. Additivity in such situations strongly implies functional and neural independence (see, e.g., Kutas & Hillyard, 1980a).

An investigation of the relationship between P600 and P300 is currently underway in our laboratory. In an initial experiment, we manipulated both anomaly type and the task relevance of the anomalies. The stimuli include three sentence types: well-formed sentences with no anomalies; sentences containing a verb that disagreed with the subject noun in number (which should elicit the P600); and well-formed sentences containing a word in uppercase letters (the type of "physical" anomaly known to elicit the P300-like effects). This allowed direct comparisons of the scalp distributions of the ERP response to each anomaly type. Additionally, the subjects' task was manipulated in a between-subjects manner. One group of subjects was asked to make "sentence acceptability" judgments following each sentence. Subjects were explicitly told that an anomaly of any type was sufficient to render the sentence "unacceptable"; hence, both types of anomalies were directly task relevant. A second group of subjects was asked to passively read each sentence. Presumably, the anomalies in this condition were less task relevant than in the first condition.

ERPs to the critical words in each sentence type are shown in Fig. 2.5A (acceptability-judgment condition) and 2.5B (passive-reading condition). In both conditions, as expected, the uppercase words elicited a large positive wave with an onset around 200 ms and with a peak amplitude around 400 ms (similar to previous reports of the P300 component; cf. Donchin, 1981). The agreement violations elicited a positivity with an onset around 500 ms and a peak amplitude around 700 ms (the P600 component). Although the amplitudes, onsets, and peak latencies of these effects differed, the scalp distributions of these effects were very similar. Furthermore, the manipulation of task relevance had a similar

[4]Recent work has indicated that there might be many "P300-like" components, each with an independent neural source (cf. Johnson, Jr., 1993). Therefore, a better way to phrase this question might be to ask whether the brain response to anomalies that involve formal, rule-governed aspects of language is distinct from the response to anomalies that do not. In the present discourse, one can think of the terms "P600" and "P300" as shorthand for these two categories of anomaly.

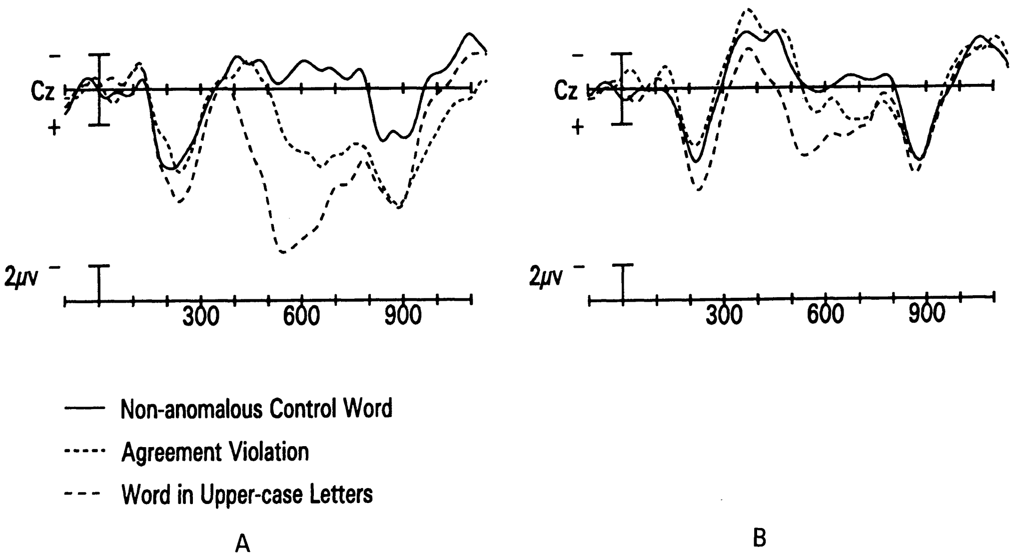

FIG. 2.5. ERPs to uppercase words (long dashes), agreement violations (short dashes), and nonanomalous control words (solid line). (A) ERPs elicited during the acceptability-judgment task; (B) ERPs elicited during the passive-reading task.

influence on the response to both anomaly types. The amplitude of both responses was greatly reduced in the passive-reading condition, relative to the acceptability-judgment condition.

These findings seem to indicate that the P600 might indeed be another manifestation of the P300 family of positivities. However, preliminary results from a second experiment lead us to suspect that the contrary claim is correct. In this experiment, we presented the sentences described earlier, plus a fourth sentence type that contained a "doubly anomalous" word (i.e., a verb that disagrees with its subject in number and that is also in uppercase letters). The goal was to determine whether the P300 and P600 are additive. We evaluated this by comparing the observed waveform elicited by the doubly anomalous words to a composite waveform created by adding the ERP response to agreement violations and uppercase words (when these anomalies were presented independently) to the waveform elicited by the nonanomalous control sentences. The composite waveform was remarkably similar to (and did not differ significantly from) the observed waveform (Fig. 2.6). We believe that this striking preliminary result represents compelling evidence that the P300 and P600 components have additive effects. The clear implication of this result is that the P300 and P600 are indeed independent. Of course, this claim in turn does not *necessarily* imply that the P600 is language specific.

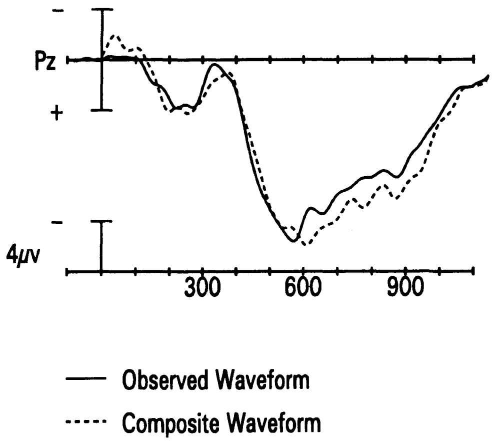

FIG. 2.6. ERPs to words that were in uppercase letters and that indicated an agreement violation (solid line) and composite waveform (dashed line).

Caveats and Complications

Perhaps the most important caveat attached to the previous work is the artificiality introduced by the relatively slow word-by-word visual presentation of sentences. Simply put, people do not usually comprehend language in this manner. Therefore, one must ask whether these results will generalize to more natural comprehension situations. Although a complete answer to this question is not yet available, early indications are encouraging in this regard. In particular, Osterhout and Holcomb (1993) replicated their earlier study (Osterhout & Holcomb, 1992), replacing word-by-word visual presentation with sentences presented as continuous, natural speech. Violations of verb subcategorization and phrase-structure constraints elicited a positive-going wave similar to that elicited during comprehension of the visual sentences, indicating that mode of presentation is not a critical factor.

A second issue concerns a possible alternative interpretation of these ERP effects. In the work reviewed earlier, large N400s have been elicited by content words (nouns and verbs), whereas the P600 has been elicited by function words (e.g., infinitival markers and auxiliary verbs).[5] Hence, the possibility exists that the N400 and P600 are elicited as a function of word class, rather than anomaly type; i.e., anomaly type has been confounded with word class). Closed class words serve primarily as vehicles of phrasal construction, whereas open class words are primarily agents of reference. A considerable amount of evidence from diverse lines of investigation suggests that these two classes are treated differently during comprehension, and might indeed involve the operation of neurally

[5]Note that in the Osterhout and Mobley (1993) study discussed above, in some sentences the critical words were (open class) verbs. However, the critical comparisons involved erroneous plurality markers, which are often considered to be closed class morphemes.

distinct systems (e.g., Bradley, Garrett, & Zurif, 1980; Friederici, 1983; Neville, 1992). Perhaps a syntactic anomaly engenders a processing response that is similar to that engendered by a semantic anomaly, but the similarity in responses is obscured by differences in the responses to tokens of the two word classes. To investigate this possibility, we recently presented garden-path sentences similar to Sentence (17):

(17) The boat sailed down the river *sank* during the storm.

Considerable prior work suggests that subjects will initially attempt a simple-active interpretation of the sentence, rather than the appropriate reduced-relative clause analysis (e.g., Bever, 1970; Frazier & Rayner, 1982). Under a simple-active analysis, the sentence becomes ungrammatical at the verb *sank,* which is a content word. If the N400 and P600 are elicited as a function of anomaly type, this word should elicit the P600. Conversely, if these effects are elicited as a function of word class, this word should elicit a large N400 effect.

Grand-average ERPs to the critical words (and to matched control words in sentences such as "The boat sailed down the river and *sank* during the storm") are shown in Fig. 2.7A. Inspection of this figure seems to indicate that the anomaly elicited a *biphasic response,* i.e., *both* an N400 and P600 effect within the same epoch. However, inspection of individual waveforms revealed that *no individual subject showed a biphasic response* to these anomalies. Rather, the majority of subjects showed a *monophasic* response to the anomalous word. In the majority of these subjects (nine subjects), the anomaly elicited a very clear P600 effect (Fig. 2.7B); in four subjects, the anomaly elicited an enhanced N400 (Fig. 2.7C); and in two subjects, the response to the anomaly did not differ from that of controls.[6] When averaged together, these monophasic responses took on the appearance of a biphasic response.

These findings are quite disturbing from a methodological point of view because averaging over subjects is a standard procedure for ERP researchers. At the minimum, these findings suggest that researchers should examine individual subject averages, rather than relying exclusively on averages over subject. A corollary of this is that researchers should present sufficient numbers of items in each condition so that the signal-to-noise ratio for each subject is sufficient to allow inspection of individual subject data.

From a theoretical perspective, the apparent existence of individual differences in subjects' responses to this type of anomaly raises some fascinating questions. For example, do these differences among subjects reflect differences in linguistic processes and capacities (e.g., differences in the grammatical com-

[6]It is not the case that the subgroup of subjects who elicited an N400 response to these anomalies would not produce a P600 response to any anomaly. Embedded within the list were agreement anomalies, which elicited a P600-like response in most subjects.