2

Preliminaries

In the present chapter, we present a framework and methodology for the empirical study of psychonarratology and discuss some of the epistemological issues that form the background for this kind of research. Following a discussion of the domain of psychonarratology, we elaborate on four aspects of the methodology that are central to its study. First, we discuss the distinction between features and constructions introduced in Chapter 1 and describe criteria for developing useful textual features. Second, the term "statistical reader" is introduced; this term describes an approach in which aggregate measures of groups of individuals are used to provide insights into the general characteristics of populations of readers. Third, we sketch some of the epistemological assumptions involved in conducting empirical research in psychonarratology and outline the theoretical goals. Fourth, we argue that the strongest inferences about reading processes can only be obtained by conducting "textual experiments" in which the text is manipulated and concomitant changes in readers' responses are observed. Together, these notions provide a foundation for the empirical investigation of the problems of psychonarratology.

This chapter was designed in part to address the needs, interests, and concerns of literary scholars who may be intrigued by the empirical study of literary response but lack the confidence to pursue it on their own. In particular, we have endeavored to outline the fundamentals of empirical research without recourse to specialized

knowledge or vocabulary. As we discuss in Chapter 1, the very word "statistics" may have unpleasant connotations for literary scholars. We hope that the following sections will demonstrate not only that these connotations are ill-deserved but, more importantly, that statistical concepts are relevant to literary studies. The actual mechanics of using statistical formulas are generally not important to understanding the theories and evidence we present in this book. However, the elements of these calculations are described in the Appendix, and this material could serve as a resource for those individuals who wish to proceed and try their hand at this kind of research. To those individuals, we would also reiterate our belief that the epistemological and conceptual underpinnings provide the greatest barrier to conducting empirical research and that the process of manipulating the numbers is relatively simple.

The Domain of Psychonarratology

The goal of psychonarratology is to understand the psychological processing of narrative form. As described in Chapter 1, one of the major advances in recent work on narratology is an agreement that what readers do with the text is crucial for an understanding of narratives and how they function. In our view, there is a common supposition that the process of reading must be considered to understand the way in which narratives function. However, there is much less consensus on how to proceed from that starting point. Our view is that narratologists have been unable to pursue that agenda effectively to date because the important issues are essentially empirical questions that can only be answered by systematically observing readers as they read. Inevitably, the answer to the question "What do readers do with the text?" will be complex: What readers do will vary with the characteristics of the individual reader, the nature of the text, and the context in which the reading takes place. This means that what is required is a large body of empirical evidence on how these variables operate, how they interact, and how they combine to determine readers' processing. Although this is a daunting project, it is essentially tractable and can be addressed with established methodologies and experimental paradigms.

In fact, in psychology and psycholinguistics, there is already a great deal of evidence and theorizing on how readers process text. This work provides a solid foundation for discussing a range of problems, including the mechanics of reading (i.e., how the eye is moved over a text and how words are recognized), the identification of sentence structure and local textual coherence, the role of working memory in on-line processing, and the variables involved in remembering and learning from text. Although important, this work does not bear on the processing of narrative form directly and literary processing generally. There are also substantial areas of discourse processing, some of which are briefly reviewed in Chapter 1, that are more directly relevant to narrative. Problems such as the structure and use of referential representations, the generation of inferences, and the role of causal structure all form a background for the present work. However, since the pioneering work of Rumelhart (1975) and others on story grammars, relatively little of this work has been concerned with narrative form per se, and the history of scholarship in narratology and literary studies has had a minimal impact. In a sense, psychonarratology can be viewed as an extension of discourse processing that is concerned precisely with narrative and that builds on the knowledge and insight gained in decades of work in narratology and allied fields of study. As such, it depends on epistemological underpinnings that are common in psychological research but that are rarely discussed explicitly. In this chapter, we attempt to make some of these foundational notions clear.

The field of narrative processing is very broad and includes such things as how readers process whole texts, different genres, works from different periods, as well as how any of these is affected by culture, class, gender, and so on. In the present work, we focus primarily on how a relatively circumscribed collection of readers process a limited set of textual features. Consequently, the work and theoretical analyses we present here comprise only a starting point. This seems appropriate given the limited empirical evidence we have in this domain: It is necessary to start at a basic level before tackling deeper and wider problems. For example, the present work may be a building block for other inquiries in cultural studies or other allied disciplines.

Features and Constructions

Applying Features and Constructions

A distinction between text features and reader constructions eliminates some critical obstacles to the scientific study of narrative. For example, as we elaborate in Chapters 3 and 5, narratologists have devoted effort to defining the nature and locus of the narrator and characters in narrative. Is the narrator "in" the text, a property of the communicative transmission of information from author to reader, or some generalization of reader response? In our view, questions such as these have been difficult or controversial to answer because of a failure to carefully distinguish between textual features (i.e., objective and identifiable characteristics of the text) and reader constructions (i.e., subjective and variable mental processes). In fact, we would argue that the narrator is simultaneously both "in" the text and "in" the reader: There are identifiable narratorial features in the text (such as speech styles, proper nouns, deictics, and propositions attributing thought and action to agents) that provide cues and information to the reader; readers in turn commonly use such cues to construct a representation of the individual who might be producing that narrative, that is, the narrator (cf. Dixon & Bortolussi, 2001b). Some hypotheses concerning the content of this representation are taken up in Chapter 3.

We believe that a distinction between features and constructions is essential to bring the right tools to bear on different aspects of the problem of understanding narrative processing. A reasonable model of how this can work can be found in the field of psycholinguistics. Psycholinguistics involves the use of empirical methods from cognitive psychology to investigate how features of language are processed by language comprehenders. In many cases, formal linguistics provides the conceptual distinctions and interesting variables for this investigation, while cognitive psychology contributes the empirical methods and interpretive tools. As a small example, consider that formal linguistic analyses indicate that the structure of sentences with reduced sentential complements are momentarily ambiguous as they are being read. Following this kind of analysis, psycholinguists have used experimental techniques such as the measurement

of the duration of eye fixations to understand how readers process that ambiguity (e.g., Frazier & Rayner, 1982). By analogy, we use the term "psychonarratology" to refer to the interdisciplinary investigation of narratorial processing. In our view, psychonarratology should consist of the empirical investigation of how the distinctions and variables identified by narratologists are processed and how they affect readers.

The Nature of Textual Features

At the heart of the successful development of psychonarratology is the identification of textual features. Here we present some criteria for what a valuable textual feature should be. We suggest that features should be objective, precise, stable, relevant, and tractable. A discussion of each of these criteria follows.

First, a textual feature must be *objective*. Objective means that the definition of the feature is public, clear, and understandable. It is not useful, for example, to describe a text as having "a sense of intimacy" because it is unlikely to be clear to many what that means exactly, and there would probably be little agreement as to which of a collection of texts have a sense of intimacy and which do not. In contrast, the presence of a first-person narrator is a much more objective feature; it is likely that most researchers would be able to agree on which of a set of texts were written in the first-person mode. Objectivity may be thought of as a continuum, and the definitions of features may evolve and become more objective. For example, one may start with a subjective insight that a text is "intimate"; subsequently, one may define an intimate text in terms of other properties of the text, such as first-person narration, the presence of substantial amounts of self-reflection, and personal judgments of other characters. As each of the elements of this definition is made more and more explicit and detailed, the definition of what it means to be an "intimate" text becomes more objective. One way of thinking about objectivity is to conceive of a test procedure that can be carried out in a more-or-less mechanical manner that would distinguish the presence of the feature from its absence. If there is such a procedure, and if it can be communicated to other knowledgeable individuals, then the feature is objective. Difficulty in conceiving of such a test suggests

that the feature may not be sufficiently objective or explicit. As a proposed feature becomes more objective, the nature and content of the corresponding test becomes more clear and detailed.

Related to the notion of objectivity is that of *precision*. With a perfectly precise definition of a textual feature, it would always be clear when a text has a feature and when it does not. Imprecise features admit of a substantial number of ambiguous cases in which it is unclear whether the feature is present. Although related, precision can be distinguished from objectivity. For example, the feature of having a substantial amount of quoted speech is perfectly objective: The testing procedure might consist of simply counting the percentage of words that appear in quotes. However, the definition, as it stands, is imprecise because it is not clear how large a percentage is needed to count as "substantial." It might serve, for example, to distinguish a text in which 80 percent of the words are quoted from one in which there is no direct speech at all, but what about a text that has 60 percent quoted speech? Or 30 percent? A lack of precision is usually not a serious problem as long as the definition can, in principle, be extended or elaborated to handle those cases that lie in the gray area. One might, for example, say that greater than 75 percent quoted speech is "predominant," greater than 40 percent is "substantial," greater than 20 percent is "significant," and less than that is "minor." In this way, the definition of the feature becomes more articulated and precise as the need for such precision arises.

A third criterion for a textual feature is that it be *stable*. By stability we mean that it is an enduring property of the text and does not vary with the reader or the reading situation. For example, "evokes a feeling of sadness" is not a stable feature because it might easily be the case that some readers do not experience this emotion while other do. It would be better instead to describe the text as having a depressing topic, or describing sorrowful moods in characters, or one in which the narrator is depicted as depressed, as the case may be. Indeed, "evokes a feeling of sadness" is a description of a construction on the part of the reader, and, as we argued previously, it is crucial to distinguish reader constructions from textual features. However, as a stepping stone to more objective and stable definitions of features, we believe that it would not be remiss to include a certain amount of potentially ill-defined interpretation as part of the definition of a

feature. For example, one might begin by describing the narrator's voice as having a dejected tone, even though one has only limited insight into the textual details that conspire to produce that tone. Eventually, one would hope that such a description would develop sufficiently so that such global interpretation would not be necessary in the definition. It may be defensible in the interim, though, as long as informed readers can generally agree on that interpretation. In this sense, the definition could be relatively stable over different readers, even though it may not be completely objective.

A fourth criterion for a useful textual feature is that it be *relevant*. There are a plethora of possible features that one might define for texts that are perfectly objective, precise, and stable. One could count the number of words that have *e*'s in them, or measure the average number of words between occurrences of compound sentences, for example. However, features such as these would not be useful in developing a theory of psychonarratology because they are unlikely to have any impact on readers' processes. It is, of course, ultimately an empirical question as to whether a particular feature is relevant. For example, whether or not a text has a substantial amount of quoted speech might affect how readers process a text, but it might not. In general, one cannot know whether a feature is relevant until suitable evidence has been gathered. Our goal in stating relevance as a criterion is to indicate that, in describing texts, one should begin by seeking features that are important for readers, and our sense is that a little introspection and common sense can go a long in way in delimiting the initial search for relevant features. In contrast, logical coherence, elegance, or parsimony are of little value in a system of features if those features have little impact on most readers.

Finally, a textual feature should be tractable. Our goal is to investigate how textual features are related to reader constructions, and that inevitably involves the conduct of research in which features are measured or manipulated. Tractability of the features means that research can be carried out in a manner that is efficient and cost-effective. For example, one might categorize a novel's plot as typical or atypical by surveying a large sample of popular fiction, rating the similarity between the novel and each of the popular works, weighting that rating by the number of copies sold, and then calculating the average. Although such a measure might easily have an impact on

the reader (mediated, we expect, by reader expectations), it would be expensive in terms of time and money to construct, and we would regard it as not very tractable. Although it is occasionally important in science to go to great lengths to measure aspects of the world that are otherwise unknown, our view is that such circumstances are unlikely to arise in the immediate development of psychonarratology. In this instance, plots might be classified as typical or atypical purely on the basis of the intuition of the researcher. Although such a measure is not completely objective, suffers from a certain degree of imprecision, and may not be entirely stable, it is simple and inexpensive, and probably a good place to start. More precise and stable measures can be developed as needed.

In many cases, definitions of textual features may be imperfect and may contain a variety of deficiencies with respect to these criteria. This is not, in itself, a fatal flaw. The essential aspect of this enterprise is simply that investigators agree on the criteria. Then, as deficiencies become clear, new and better definitions can be developed as the need arises. One need not have a perfect, pristine set of carpentry tools to build a house; one simply starts to work with the tools at hand and then obtains new ones if the old tools prove inadequate.

Measuring Reader Constructions

A second component of our methodological framework is the measurement of reader constructions. There is a wide range of techniques for measuring such constructions; the approaches we have used in our research represent only a tiny fragment. It is sometimes suggested that the empirical investigation of literary processing requires the development of new, sophisticated measurement methodologies that are specially suited to the domain (Schmidt, 1981). Sophisticated tracking of eye movements during reading or the detailed analysis of verbal protocols might be suggested, for example. Although the intelligent use of such methods can be very insightful, they can also be extraordinarily expensive and time-consuming. Further, we disagree that such methods are always necessary. Instead, we believe that the most mundane methods imaginable (such as simply asking readers direct questions concerning the text) are often sufficient when they are combined with a careful delineation of features and constructions

and when the investigation is conducted with due regard for the need to make strong inferences concerning their relationship. Nevertheless, there are likely to be many circumstances in which one may wish to consider other ways of assessing reader constructions, and here we mention what we believe are some important criteria for such measurements. In particular, we argue that the measurement of reader constructions should be direct, replicable, and concise.

A measurement is direct when what is actually recorded corresponds closely to the construction of interest. Any empirical investigation of mental processes involves a significant degree of what can be termed "operationalization." For example, one may wish to measure the extent to which a reader sees the narrator as similar to him or herself. One cannot see inside the reader's mind to make this measurement, and the researcher must instead be satisfied with a somewhat more indirect approach, such as asking readers to rate their similarity to the narrator. In this instance, readers' rating response is an operationalization of the construction of interest, readers' perceived similarity; one would say that perceived similarity has been operationalized as a rating response. With a little care, the two measurements are likely to be closely related so that the operationalized measure provides a good index of the reader construction. However, even in this simple example, they need not be precisely the same; some readers may be motivated for some reason to lie in their questionnaire response, or they may misinterpret the questions in some way. Having a measurement with minimal operationalization means that the link between the reader constructions and what is actually measured is relatively short and direct and that the likelihood of such extraneous effects is small. In the vocabulary of measurement theory, a minimal operationalization has "face validity."

Replicability refers to the ability of other researchers to conduct a similar investigation and measure the same reader construction. For example, if readers fill out a questionnaire under controlled circumstances, those measurements are usually quite replicable because another researcher need only know the precise wording of the questions to conduct virtually the same investigation. In contrast, if a researcher simply reports that readers failed to appreciate the import of a given work, such a measurement is likely to be difficult to replicate, even if the report is correct. The problem in this case is that it is difficult

to ascertain what precisely might be meant by "failed to appreciate" and that the import of a work is potentially open to interpretation; as a consequence, other researchers might easily not find that reported result. In general, replicability is greater when the measurement instrument is described objectively and when all the relevant details of the measurement procedure are explicit.

Conciseness in a measurement instrument means that the results of the measurement can be communicated in an efficient manner. Unabridged verbal protocols, for example, might easily reflect reader constructions in a manner that is direct and replicable. However, they do not provide, by themselves, a particularly useful measurement instrument because they cannot be easily communicated. To be useful, protocols typically need to be summarized and scored in some manner. Only then would they be appropriately concise and useful as an index of reader constructions. However, the method of scoring and summarization must properly be thought of as part of the measurement. As such, they must be suitably direct (i.e., the summarization itself should reflect reader constructions) and replicable (i.e., other researchers should be able to summarize the same protocols and get similar summaries). In contrast, numerical measurements almost always provide concise measurements because the methods for summarizing and describing a body of numerical measurements are well established, and under most circumstances one may reasonably assume that standard descriptive statistics provide a direct and replicable description of the original data set.

The Statistical Reader

One of the greatest challenges for empirical approaches to literary reception is the need to reconcile two seemingly contradictory assumptions concerning the nature of the reader. On one hand, there is the recognition that the response to and interpretation of a literary work is not homogeneous within any group of readers, and that, instead, there is a range of responses and reactions that must be explained. On the other hand, there is the assumption that literary processing is not entirely idiosyncratic, and that the processes used by readers are amenable to scientific investigation and description. We have developed an explicit formulation of this problem in our notion

of the "statistical reader" (Dixon et al., 1993). This development derives largely from some basic tenets of probability and statistics; hence the term "statistical." It may be apparent that the view we are proposing here is not new, but rather a restatement of many of the implicit assumptions involved in standard data analysis procedures. However, there are crucial advantages in making these notions explicit for the development of the present methodological framework.

Central to this issue is the question of how variability in reader constructions is interpreted. We do not believe that such variability should be interpreted as random noise or error; each reader's constructions must be viewed as valid and appropriate given that reader's knowledge, background, goals, and personality. At the same time, we do not believe that the study of reading processes should be limited to simply describing specific and idiosyncratic interpretations of a text, as some radical reader-response critics have done (e.g., Holland, 1975, 1980). Instead, we view the task of a science of psychonarratology as the description of the properties of the processing that are common to groups of readers in aggregate. Moreover, we posit that scientific claims about literary processing and interpretation cannot be made in the abstract, but only with respect to some explicitly defined population of readers. However, reader populations are unlikely to be uniform and monolithic with respect to interesting aspects of narrative processing. Instead, we view any given reader population as consisting of a complex collection of overlapping and nested groups, each with potentially distinct reading processes. In this respect, we believe that the concept of the statistical reader is an advance over the vague, informal, a-historical, and essentialist notions of the reader that have often prevailed in literary studies.

The notion of the statistical reader is thus based on two powerful concepts: population and measurement distributions. We first present a relatively formal description of these concepts in their simplest form. We then describe how these notions form the basis for a description of variables and relationships.

Population

A population is a collection of individuals about which interesting claims might be made. For example, a population might be all

undergraduate students in the United States and Canada; another might be skilled readers of English; still another might be literary critics who have published papers on Borges. We propose that any scientific claim about reading, processing, and interpretation must be prefaced with a description of the population to which the claim applies. In other words, scientific claims cannot be justified in the abstract, but only with respect to a particular domain of reference. However, there is no restriction on size, scope, or criteria for defining a population, and there may be any number of populations about which interesting claims might be made. In particular, literary processing in one population might proceed in one way, and in an interestingly different way in another. Moreover, the multitude of interesting populations may be related in complex ways: They may be nested, so that one population is a subset of another; they may be disjoint, so that no individuals in one population are in another; or they may overlap, so that some individuals in one population are also contained in another. Sorting out the facts that apply to these various populations is a difficult but empirically tractable problem. However, its solution requires that the populations be specified at the outset.

Measurement Distributions

A measurement is anything that one can assess or evaluate about an individual, his or her behavior, or his or her situation. Measurements are objective in the sense that it is possible to explicitly describe the procedure by which the measurement was collected; this procedure is the measurement instrument. Although measurements are often quantitative, they need not be. Measurements might be disjoint categories (e.g., good reader or bad reader), lists (e.g., the books read this year), or even open-ended verbalizations. The set of measurements that might, in principle, be collected from an entire population is referred to as a measurement distribution. The notion of a measurement distribution is crucial because it admits the possibility that populations are heterogeneous with respect to any given measurement. In particular, we assume that measurements inevitably vary and are generally different for different individuals or on different occasions. However, despite this variability, a measurement distribution can

be described. Two classes of descriptions are common: Descriptions of *central tendency* (e.g., the mean or average) indicate what measurements are likely to be found in the population; descriptions of *variability* (e.g., the standard deviation) index the range of possible measurements. Although measurements are generally collected from some small sample of individuals, those measurements may provide a substantial amount of information about the measurement distribution for the population. In fact, well-established procedures are available for generating precise mathematical descriptions of the distribution from relatively small samples selected from that distribution. For example, if measurements are randomly selected from the measurement distribution, the mean of the sample provides an estimate of the mean of the distribution, and the precision of that estimate increases in a lawful manner with the size of the sample. Similarly, the variability of the distribution (described, for example, by the standard deviation) can be estimated by the standard deviation of the sample. That is, one may infer, with some degree of confidence, properties of the measurement distribution from a restricted sample of measurements. Thus, an empirical approach, in which measurements are assumed to be sampled from a measurement distribution, provides a mechanism for describing not only how individuals and their reactions tend to be the same but also how they differ. Both are required for a clear understanding of readers' constructions.

Measurements are also made of texts. For example, one might identify whether a text has a particular feature or how much of a feature a text has. However, because texts are not (usually) sampled from some population, there are typically no corresponding notions of central tendency or variability. (It is worth noting, however, that this view is not universally held. For example, following Clark (1973), it is common in psychology to treat words, sentences, and other verbal materials statistically as if they were randomly sampled from a population of comparable materials. See Wike and Church (1976) and Clark et al. (1976) (for some of the debate on this practice). Texts do vary, though, from one to another, and this is a critical ingredient in the identification of relationships between textual features and reader constructions.

Variables and Relationships

Describing distributions and their characteristics is the first step in identifying variables and relationships. For our purposes, we define a variable as a set of measurements that can be indexed in a systematic manner. For example, suppose we are interested in whether readers believe the narrator of a story is male or female. The measurement instrument in this case might be a rating response in which readers select a number on a scale with, for example, 1 corresponding to "certain to be male" and 7 corresponding to "certain to be female." The population might be University of Alberta undergraduates. The measurement distribution would correspond to all the ratings that might be produced by the individuals in that population. In this situation, a variable would be generated by collecting several sets of related measurements. For example, ratings might be collected by having readers report their perception of the narrator in several different works; the index in this case would be those works. For each value of the index (i.e., for each different work), there would be a corresponding measurement distribution.

Variables can be classified in terms of the nature of the measurements. In the previous example, the measurements are of reader constructions; consequently, the generated variable can be referred to as a *construction* variable. When the measurements are of texts or textual features, we may refer to a *text* variable. For instance, one might classify different works on the basis of whether the main character is male or female. As mentioned earlier, text measurements typically do not involve sampling from a distribution. Nevertheless, such measurements can be indexed and can be used to generate variables. Thus, having a male or female main character can be a variable, indexed by different works. An important third class of variables pertains to different types of reader populations and can be referred to as *reader* variables. For example, one population of readers might be undergraduates at the University of Alberta; another might be graduate students in literary studies at the University of Alberta; still another might be literature instructors at the University of Alberta. The measurement distributions in this case are indexed by the type of reader (corresponding roughly to level of expertise). Although reader

variables are, in principle, central to the study of psychonarratology, we will have relatively little to say about them in the present work.

Inferences in empirical investigations are possible only when variables are related to one another. A relationship in this sense means that there is concomitant variation between two variables that are indexed in the same way; in other words, the variables covary. For example, it might be the case that works that have a female protagonist are more likely to be viewed as having a female narrator than those that have a male protagonist. In this instance, having a female or male protagonist is a text variable indexed over works; viewing the narrator as female or male is a construction variable also indexed over works. The variables are related because, as one goes from one work to another, the measurements tend to change together: male protagonist and the perception of a male narrator in one case, female protagonist and the perception of a female narrator in the next, and so on. In the research that is presented here, most of the relationships will be of exactly this type: Textual variables, indexed over different texts, will be related to construction variables, also indexed over texts. However, other relationships are important as well, such as those between construction variables and reader variables.

It is important to distinguish *meaningful* relationships from trivial ones. A meaningful relationship is one that bears on interesting theoretical considerations. Typically, this means that the covariation is of a magnitude that is sufficient to distinguish one theoretical view from another. One could identify a great many possible relationships in the processing of narrative that, although real, are of little importance in developing a theory of psychonarratology. For example, it could be the case that novels that begin with a question lead to somewhat more active processing on average than those that begin with a declarative sentence; perhaps further, this small difference in processing has a miniscule persisting effect over the course of the book, so that readers end up with a very slightly more elaborate representation of the story world by the novel's end. If this chain of reasoning is correct, it implies that there should be some (very small) covariation between the form of the initial sentence and number of details in readers' representation of the story world. Such covariation would be exceedingly difficult to detect with an empirical study. Moreover, even if it were detectable, the magnitude of the effect is

likely to be so miniscule that there would little point in developing one's theoretical analysis of narrative processing in order to incorporate such influences. In other words, the potential relationship is trivial. Although this example is clear, the magnitude of the covariation required to make the relationship meaningful in most situations is a matter of debate. However, our point is that the mere existence of some degree of covariation is not necessarily interesting or relevant to the development of a theory; a relationship must be meaningful as well. To be meaningful, covariation must always achieve some minimal magnitude, and anything smaller than that is simply not of concern.

Theory and Explanation

As described in Chapter 1, our goal here is to develop a framework for the theoretical and empirical investigation of the psychological processing of narrative. Although we do not have a complete theory of this domain as yet, we have some sense of what such a theory would entail. In particular, we believe that a theory consists in large part of a set of causal explanations. A causal explanation indicates that variables are related because one causes another. For example, one might observe that a narrator tends to be interpreted as sympathetic to a character when that character's thoughts are commonly presented in the narrative. A causal explanation of this observation is that the textual characteristics cause that reaction in readers. It is critical in developing such theories to distinguish the *observation* (i.e., that readers' reactions tend to co-occur with particular aspects of the narrative technique) from its *explanation* (i.e., that readers' interpretations are caused by the use of that technique). An observation is subject to validation and replication but cannot be refuted by argument or opinion; an explanation, on the other hand, must always be considered a hypothesis subject to further support or testing and can never be proved beyond doubt.

Clearly, causal explanations in psychonarratology would not allow a simple, unequivocal determination of what a reader's reaction is likely to be. Instead, they must be understood as describing tendencies or proclivities that operate in conjunction with a network of other influences. For example, although presenting a character's thoughts

in the narrative might lead the reader to see the narrator as more sympathetic to the character, it is not the only variable that might produce that reaction, and these other influences might produce the reaction even in the absence of the character's thoughts in the narrative. Similarly, there may be other factors that lead the reader to perceive the narrator as relatively unsympathetic, and these other influences may prevail even when the character's thoughts are present. An appropriate way to view this situation is to conceive of two otherwise identical situations, one in which the textual feature is present and one in which the feature is absent; if the causal explanation is correct, readers should see the narrator as relatively more sympathetic in the former. In other words, the hypothesized causal explanation is that, all other things being equal, putting the character's thoughts in the narrative should lead to a stronger perception of a sympathetic narrator.

Any given observation typically admits of a variety of causal explanations. For example, it may not be the presence of the character's thoughts per se that leads the narrator to be seen as sympathetic, but rather the fact that the character's thoughts provide a sensible and coherent explanation of the character's behavior. Alternatively, it may be the case that whenever the author intends to present a sympathetic narrator, a variety of techniques are used, including the presentation of the character's thoughts, but only a few of these are causally related to the reader's interpretation. More complex causal explanations must also be considered. For example, the relationship between the narrative technique and the reader's reaction may only obtain if readers are predisposed to reflect on the attitude of the narrator or if other aspects of the narrative make the attitude of the narrator salient. Identifying and distinguishing at least a few of these possibilities is the substance of the theory we present in this book.

In our approach, as in any line of empirical investigation, progress at developing causal explanations is of necessity piecemeal. Possible explanations may vary with the situation and characteristics and goals of the reader, and a complete causal explanation of a given reader construction is likely to entail a myriad of interactions among a wide range of textual features. Because of this potential complexity, it is impossible to test or even describe a causal explanation of any

depth given our current state of knowledge and evidence. However, we believe that such detailed causal explanations are possible in principle; they merely require a great deal of empirical and theoretical work. Given this perspective, it would be ingenuous of us to suggest that the present approach and set of hypotheses are unassailably accurate. Instead, they merely scratch the surface of the empirical work that is needed on this problem. We believe that progress is made by proposing specific, testable causal explanations and then revising and elaborating those explanations as needed. The crucial ingredient in such progress is compelling empirical evidence.

The Textual Experiment

As described earlier, causal explanations are central to a framework for psychonarratology. We anticipate that the most important causal explanations are likely to be those that explain particular reader constructions as caused by particular features of the text. For example, we might hypothesize that the use of free-indirect speech causes a reader to have a sympathetic attitude toward a character; that the narrative description of multiple hazardous outcomes creates suspense in the reader's mind; and so on. We argue that the best technique for assessing such causation is to conduct *textual experiments*, in which particular features of a text are identified and manipulated by the researcher (Dixon & Bortolussi, 1999c). In a properly designed textual experiment, several versions of a text are created that are identical except for the single, manipulated feature. Experiments of this sort have a special relationship to causal explanations and, as a consequence, should be used whenever possible to evaluate hypotheses concerning the connection between features and constructions.

Covariation and Causation

The logical basis for this claim is illustrated in Figure 2.1. Generally, causes are identified by covariation between two variables, denoted as *I* and *D*. Covariation occurs whenever the particular values of one variable tend to co-occur with particular values of the other variable. For example, suppose that *I* is the presence of a character's thoughts

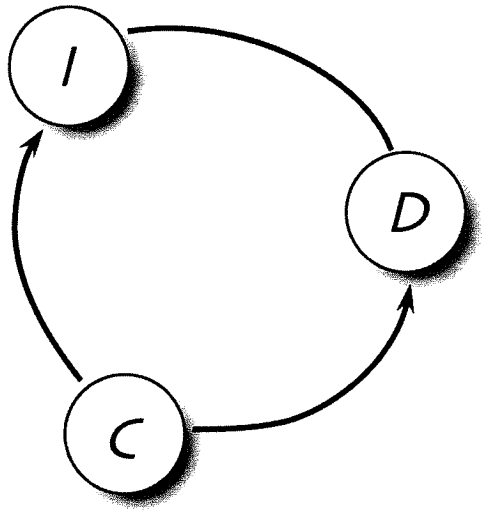

FIGURE 2.1. Covariation and confounding.

in a narrative and *D* is the reader's view of the narrator's sympathy toward a character. Covariation between *I* and *D* would occur if narratorial sympathy occurs more often with texts that include the character's thoughts than with texts that do not. Covariation of this sort is indicated in Figure 2.1 by the arc between the *I* and *D* circles.

Whenever two variables covary, there are three possible causal explanations: The first variable can cause the second, the second variable can cause the first, or some other, third, variable may cause both of these. When *I* refers to features of the text and *D* refers to a reader construction, we are usually interested in the hypothesis that *I* causes *D*. The possibility that *D* causes *I* can safely be ruled out on logical grounds: The text precedes the reader's exposure to it and cannot be affected by the reading process and the reader's constructions. However, the possibility that *I* and *D* might both be caused by a third variable is very real and compromises any possible interpretation of *I* as the cause of *D*. In the vocabulary of experimentation, the third variable is known as a confounding variable. A possible confounding variable *C* is shown in Figure 2.1, with arrowheads to indicate that it could be the cause of both *I* and *D*.

Author Confounds and Other Artifacts

In the course of drawing inferences from observed covariation, an unexpected confound may produce a relationship as an artifact; that is, the relationship may emerge simply because both variables of interest covary with a confounding variable. Although artifactual relationships can occur in any empirical investigation, they are common when variables are indexed over texts. For example, one may find that in a selection of texts, novels with a male protagonist are rated as more literary than novels with a female protagonist, and, because of this relationship, one might consider the explanation that the gender of the protagonist caused the difference in the literariness ratings. However, on further examination, it might turn out that the sample of texts included a large group of popular romance novels (in which the protagonist is typically female) and a large group of eighteenth-century literary works (in which the protagonist is more likely to be male). As a consequence, the relationship between gender and ratings could easily be an artifact of the manner in which the texts were selected. For example, the relationship is confounded with the genre of the work or its period, either of which could have a causal effect on the literariness ratings. Because of these confounds, then, the inference that gender of the protagonist has a causal effect on ratings is suspect.

Figure 2.2 illustrates why confounding is a central problem in the analysis of the effects of literary discourse. Generally, we imagine that texts are constructed by authors who have some intention concerning the aggregate effects of the narrative; such intentions may lead them to construct the text in a particular way with a particular constellation of features and characteristics. This is indicated by the variable *A* in Figure 2.2. For example, an author may want to portray a narrator who is sympathetic to a character. If the author believes that this might be done by providing the character's thoughts in the narrative, the feature might be included in the narrative. However, the author likely would also provide other features. For example, the character may express actions or thoughts that are likely to generate sympathy on the part of the reader, the narrator might express sympathetic attitudes toward the character explicitly, and so on. These other textual features are indicated by the variable *C* in Figure 2.2

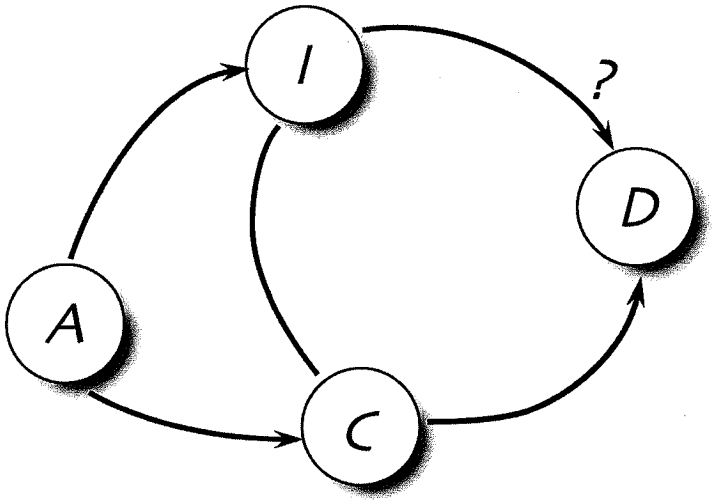

FIGURE 2.2. Author confounds.

and represent a possible confound in the relationship between the feature of character thoughts (*I*) and the reader's interpretation of the narrator (*D*). In naturally occurring narrative, the aggregate of techniques that follow from the author's intention may be the ultimate cause of both variables. Because the reader's constructions might be caused purely by these other textual features rather the feature under consideration, the covariation between *I* and *D* does not necessarily imply causation. We refer to this situation as the problem of "author confounds."

The solution to this inferential dilemma is to manipulate the text experimentally, that is, to conduct a textual experiment. For example, we, as the researchers, might construct a modified version of the author's text in which all the character's thoughts have been removed. If the reader's construction is determined purely by variables other than the character's thoughts, such a manipulation should have no effect; on the other hand, if the character's thoughts in the narrative has a causal role, one should observe some change in the reader's reactions. This situation is illustrated in Figure 2.3. In this case, only the presence or absence of character thoughts (*I*) varies across the texts used in the study; all other characteristics of the text

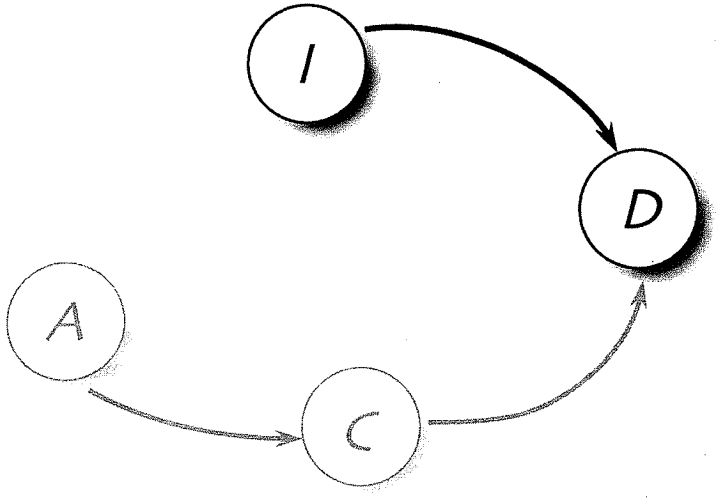

FIGURE 2.3. Unambiguous causal inference in a textual experiment.

(as well as the author's original intentions) remain the same. In textual experiments, *I* is referred to as the independent variable in the experiment and *D* is referred to as the dependent variable. In our example, any covariation between the features of the narrative and the reader's constructions can be attributed unambiguously to a causal connection between that manipulated feature and the corresponding reader construction. Such an unambiguous causal inference is an overwhelming advantage in constructing theoretical accounts of readers' processing, and it accrues uniquely to textual experiments.

Designing Textual Experiments

The logical power of textual manipulations is clear in principle. However, the technique requires that one be able to manipulate one variable without inadvertently changing other, potentially confounding variables. In turn, this requires a sophisticated theory of the text and its features and structure. Thus, the success of empirical research in psychonarratology builds on the availability of detailed and elaborate theoretical analyses such as those found in narratology. Analyses

of this sort allow one to manipulate particular features of the text confidently without introducing confounds.

Even with a sophisticated theoretical description of textual features, however, confounds might be introduced inadvertently. For example, suppose one were conducting an experiment on the effects of mental access, that is, the provision of information about a character's internal thoughts and reactions. The following text fragment would provide mental access:

The child thought his parents couldn't hear him crying. (1)

In a textual experiment, this fragment might be changed so that it does not provide mental access, and the following replacement might be used:

The child's parents couldn't hear him crying. (2)

This particular manipulation is potentially confounded: Although both text fragments describe the same situation, the first seems to suggest the possibility that the parents could in fact hear the child, despite his impression, while the second does not. In other words, the manipulation of mental access is potentially confounded with the information conveyed about the situation. Typically, such artifacts arise because of an incomplete analysis that fails to consider important variables. (Indeed, based on our analysis of represented speech and thought described in Chapter 7, we would argue that, in this particular example, the manipulation is confounded with conversational properties of the narrative.) Although undesirable, this potential confound is not a fatal flaw in the method of textual experiments, and two aspects of the methodology allow one to recover from such problems.

First, under many circumstances, an incomplete analysis that fails to identify a confound may simply lead to a weak manipulation that produces few changes in reader constructions. For example, suppose one were interested in evaluating whether mental access alters the reader's interpretation of the narrator's attitude toward the character; in particular, the hypothesis might be that eliminating mental access would lead the reader to see the narrator as less sympathetic to the character. According to this hypothesis, the narrator should be seen as less sympathetic with fragment (2) than with (1). The slightly different

information provided in these two excerpts may not compromise this hypothesis because, if anything, the situational information conveyed by (2) makes the narrator seem arguably more sympathetic to the character, not less. Thus, if results of the study confirm the hypothesis, it would be difficult to attribute those results to the confounding variable. Another way of saying this is that, although the manipulation is confounded, the expected effect of the confound is contrary to the hypothesis of interest and, as a consequence, cannot be used as an alternative explanation. If the analysis of the confound is correct (i.e., the information conveyed by (2) leads the narrator to be seen as more sympathetic), the causal inference based on the expected result remains sound.

Second, the presence of confounding variables can be readily evaluated by further experimentation. For example, if one were concerned about the possibility that the difference in the information conveyed by (1) and (2) has an effect on the dependent variable, one could simply test whether that information, by itself, has an effect. In this case, one might compare the results obtained with (1) to those obtained with the following fragment in which the new information is conveyed explicitly:

The child thought his parents couldn't hear him crying, and he was right. (3)

The data obtained from this further textual experiment would either confirm or disconfirm the causal effect of the new manipulation and would indicate clearly whether the data from the original experiment was confounded or not. A series of control experiments of this general sort can eventually lend great confidence to a particular causal inference. In general, the possibility of potential confounds is not an inherent defect in the method of textual experiments; rather, potential confounds are empirical questions that need to be addressed by (further) careful experimental investigations.

Identifying Meaningful Relationships

After the data from an empirical investigation have been collected, one needs to decide whether a meaningful relationship exists. For example, in a textual experiment, one would need to decide whether there was a relationship between the manipulation of the text

(the independent variable) and readers' constructions (the dependent variable). In this case, it is common to say that each different version of the text defines a treatment condition, and the task of identifying relationships amounts to asking whether a dependent variable varies with condition. There is a long history of debate in statistics on the appropriate way in which to solve this problem. The approach that is commonly used in psychology and many other disciplines is sometimes referred to as "null hypothesis significance testing" and derives in part from ideas proposed by Fisher (1925) and Neyman and Pearson (1928). Although this approach is widely used and taught, it suffers from a variety of well-known flaws and logical inconsistencies, not the least of which is that it fails to distinguish between meaningful relationships and trivial ones. Without going into any details, we find the arguments against null hypothesis significance testing compelling and will not report such tests here. Instead, we use an alternative method described by Dixon and O'Reilly (1999) and modeled on the approach suggested by Edwards (1992), Fisher (1955), and Goodman and Royall (1988).

The approach to identifying meaningful relationships adopted here involves reporting the evidence for different competing interpretations of the results. Typically, one wishes to compare the interpretation that the independent variable (the textual manipulation) is related to the dependent variable (the reader measurement) to an alternative interpretation that the variables are unrelated. Each of the two interpretations may match the obtained results more or less well. For example, given the interpretation that there is no relationship, any obtained difference between the conditions would be error and would count against the interpretation that the conditions are actually the same. Alternatively, given the interpretation that there is a meaningful relationship of a specified size, an obtained difference that is smaller would count against that interpretation. The evidence for one interpretation relative to the other can be succinctly captured by calculating a likelihood ratio. The likelihood ratio indicates how likely the data are given one interpretation divided by how likely the data are given the other. Large values provide support for the first interpretation, while small values less than 1 provide support for the second. As a rule of thumb, a likelihood ratio of 10 or greater (or 0.1 or smaller) constitutes clear evidence for one interpretation

over the other. (For readers familiar with null hypothesis significance testing, "clear evidence" generally implies rejecting the null hypothesis with $p < .05$; Dixon, 1998). Likelihood ratios of this sort are readily calculated by recombining the elements of more traditional statistical calculations; a summary of some of these methods is presented in the Appendix. Because likelihood ratios provide a convenient and intuitive summary of the strength of the evidence for a relationship obtained in an experiment, they are used here to document the effects we discuss.

Summary

In the present chapter, we provided some foundational concepts and methods for an empirical investigation of psychonarratology. We began with the distinction between textual features and reader constructions: Features are objectively described properties of texts, while constructions are variable mental representations generated by individual readers. The statistical reader provides a basis for assessing reader constructions empirically. Populations of readers must be explicitly identified and sampled, and the characteristics of the measurement distribution must be calculated on the basis of the sample. We argue that sound causal inferences concerning the relationship between features and constructions can only be obtained by making use of the textual experiment in which features of the text are systematically manipulated and concomitant variations in reader constructions are measured; to do otherwise admits author confounds and other artifacts. Finally, we discussed some data analytic procedures: We avoid here the common procedure of null hypothesis testing and replace it with likelihood ratio calculations.