In J. Altarriba & R. R. Heredia (Eds.), ***An Introduction to Bilingualism: Principles and Processses*****.**

2008. Mahawah, NJ: Lawrence Erlbaum.

Chapter 2
Bilingual Research Methods

Viorica Marian
Northwestern University

Bilingual Research Methods: Introduction

In 1924 the United States Congress passed what became known as the Immigration Restriction Act, a law that regulated immigration to the United States for many years and served as the basis for discriminatory immigration policies favoring immigrants from Western and Northern Europe over those from Southern and Eastern Europe. The law had an eugenic intent (eugenics refers to 'improvement of the gene pool') designed to halt the immigration of supposedly "dysgenic" groups, groups that purportedly contributed to a decline of the gene pool. The Immigration Restriction Act relied in part on data from seemingly scientific studies (Brigham, 1923; Goddard, 1914) as well as a Public Health Service project that tested the intelligence of different groups and found that some immigrant groups—for example, Italians and Eastern European Jews—scored lower, often below average, and sometimes even in the 'feebleminded' range, compared to other groups. Herrnstein and Murray, in their much-publicized 1994 book, "The Bell Curve" described these events as follows: "In the early 1920s, the chairman of the House Committee on Immigration and Naturalization appointed an 'Expert Eugenical Agent' for his committee's work, a biologist who was especially concerned about keeping up the American level of intelligence by suitable immigration policies" (p. 5).

One can just imagine how, in the years that followed, streams of immigrants lined up at Ellis Island to undergo comprehensive medical examinations, coupled with psychometric tests to assess their intellectual abilities. Never mind that many of these immigrants spoke not a word of English, while their testers did not know many of the languages that were represented among those fresh off the boats. Imagine yourself as a Ukrainian farmer, illiterate, never having taken a paper-and-pencil test in your life and speaking no English, after a long and stressful journey to a country at the other side of the world, having to take an IQ test. Is it any wonder that some groups—e.g., British, Dutch, German (whose languages were from the same Germanic family group as English and shared many common words and word roots)—had fewer problems understanding their testers and tests than other groups—e.g., Russian, Polish, Italian (whose languages differed from English a lot more)? Is it any wonder then that some groups did better than others for reasons that had nothing to do with intelligence?

The Immigration Restriction Act of 1924 was later repealed and, looking back, we can safely say that those mental tests were biased, that they did not take into account the linguistic and cultural background of the test-takers. Mental tests have come a long way since then and test makers are acutely aware of the need to create assessment tools that are linguistically and culturally sensitive. Yet, accomplishing such goals is not an easy task. To this day, mental tests seem to yield higher scores in some groups than in others (e.g., Herrnstein & Murray, 1994) and arguments about the lack of cultural and linguistic/dialectal fairness of these tests abound. Some of this cultural and linguistic biases are due to the fact that those who conduct research with linguistically and culturally diverse populations continue to be trained primarily in a context that focuses on middle-class English-speaking white populations and have a limited understanding and knowledge of what studying cognitive abilities of other groups entails. Studies focusing on linguistically and culturally diverse groups frequently yield seemingly contradictory findings, and conclusive answers to research questions remain elusive. The dearth of training on issues related to cognitive performance in linguistic and cultural minorities, together with failure to take into account relevant experimental variables, continues to pose a challenge in obtaining a clear picture of cognitive abilities in diverse populations. There is also the risk of inappropriately driving public policy, for instance, on issues related to raising bilingual children or to bilingualism in the classroom.

In this chapter, we take you through the steps necessary to conduct a research project with bilinguals, multilinguals, or second- and foreign-language learners. We discuss some of the issues in designing a study, selecting participants, putting together materials, collecting and analyzing data, and then disseminating the findings among an audience of peers. We consider strengths and weaknesses of different approaches, and discuss how to avoid the most common pitfalls in conducting bilingualism research and in interpreting the findings of already existing studies. The first part of the chapter introduces key terminology and concepts necessary to embark upon a research project. The second part of the chapter samples research areas that fall under the umbrella of bilingualism and illustrates how methodological differences and limitations can influence findings. The final part of the chapter considers specific methodological aspects in conducting a study with bilinguals. Sample questions and research projects, as well as resources for further information, are included at the end.

This chapter is intended for advanced undergraduate and graduate students, and for anyone new to research with linguistically diverse populations. Most frequently, these researchers find themselves in the fields of psychology, linguistics, communication sciences and disorders, or education, but can work in other disciplines as well (e.g., anthropology, neuroscience, etc.). Though one chapter alone is not sufficient to provide comprehensive training in such a complex area, it can serve as a starting point for those who are interested in bilingualism and want to ensure that they avoid the most common mistakes along the way.

Designing a Research Project with Bilinguals

In this part of the chapter, we introduce some of the key concepts necessary for familiarity with both the vocabulary used in research and the basic procedures in running a study. If you have never taken a research methods course before, much of this information will be new. If you are already familiar with the basics of research design, this will serve as a refresher tailored specifically toward research with bilinguals and multilinguals.

Observational and Experimental Studies

Research with bilinguals usually focuses on understanding cognitive, linguistic, and behavioral aspects of bilingualism. One way to accomplish this is to observe human cognitive and behavioral performance in natural settings, record such performance, and describe it for scientific understanding. This is usually known as *naturalistic observation* or *descriptive research*, because it describes naturally observed phenomena, instead of experimentally controlled or manipulated ones. An example of observational research with bilinguals may be observing a bilingual child on a playground and writing down the words the child uses in each language. Naturalistic observation is also sometimes referred to as *correlational research* when the focus of the study is on establishing a relationship between two or more variables. For example, one may find that the larger the vocabulary in a bilingual's second language, the higher his/her score on an intelligence test. You may have already heard the statement "correlation does not imply causation." In correlational studies, one is unable to make causal judgments about the effect of one variable on the other. In the case of the relationship between vocabulary size and intelligence, the only conclusion that can be reached is that the two variables are related. Correlational research is especially useful when it is not possible to manipulate a variable experimentally (e.g., due to ethical reasons, or because the event took place in the past) and the researcher is limited to recording behaviors as they occur naturally and then analyzing the collected data. Note that correlational research, while sometimes relying on actual statistical correlations such as Pearson R, is not limited to them in data analyses. That is, there is a difference between correlational research as a methodology and statistical correlation as a tool for data analysis. Other statistical analyses (such as t-tests and Analyses of Variance) can also be used in correlational studies, as long as the data that are being analyzed have been collected using the observational method, without any active manipulation of any variables on the part of the experimenter.

Alternatively, one may design an experiment and look at how changing variables influences cognitive and behavioral performance. This is known as *experimental research*. The distinguishing feature between experimental research and observational research is whether or not the experimenter is able to manipulate variables experimentally or is limited to measuring them as they occur naturally. An example of experimental research with bilinguals may consist of asking bilinguals to label pictures in either their first or their second language and comparing reaction times in this picture-naming task across the two languages. Experimental research makes it possible to control variables (such as word frequencies or word lengths in the two languages) and in general provides greater control over the behavioral and cognitive processes of interest. It makes hypothesis testing easier and allows one to draw causative inferences, that is, it allows one to establish some type of cause and effect. However, experimental research is not always feasible, practical, or ethical. For example, if one were interested in studying bilinguals' flashbulb memories—memories of dramatic public events such as a presidential assassination or a great disaster—across the two languages, one could not create such memories experimentally and would have to use the naturalistic approach by conducting first- and second-language interviews about, for example, bilinguals' memory for the 9/11 attacks on the World Trade Center towers and the Pentagon.

Longitudinal and Cross-sectional Research

Longitudinal studies are studies that follow experimental participants over a period of time, be it months, years, or decades. In longitudinal studies, performance at Time 1 is usually compared to performance of the same individual or group of individuals at Time 2. This is different from *cross-sectional* research, in which different individuals or groups of individuals are compared to each other at the same point in time. For example, if one was interested in measuring first-language (L1) and second-language (L2) vocabulary in children at ages 1, 2, and 3 years, one could go about collecting data in two ways. The first, longitudinal approach, would be to measure vocabulary size in the same group of twenty children over time, testing them at age 12 months, 24 months, and 36 months. The second, cross-sectional approach, would be to measure vocabulary size of three different groups of children, one group of twenty 12-month-olds, one group of twenty 24-month-olds, and one group of twenty 36-month-olds at about the same point in time. The advantage of longitudinal research is that it follows the same group under different conditions, thereby minimizing between-group differences (such as socio-economic status, for example) that may influence the findings. Another advantage is that it allows for a smaller sample size of participants and is therefore usually the preferred choice when studying unique groups, such as speakers of an endangered language or bilingual children with Specific Language Impairment. The disadvantage of longitudinal research is that it usually has higher attrition rates, with more participants dropping out of the study, moving away, or undergoing a life change that makes it impossible to continue with the experiment. Moreover, longitudinal research can take a long time, making it less-than-ideal for those researchers who have to work within time constraints, such as undergraduate and graduate students who will graduate before their infant participants enter college. The advantages and disadvantages of cross-sectional research are precisely the opposite to those of longitudinal research. On the up-side, cross-sectional studies take less time to run and in that way are the more practical choice. On the down-side, there are more differences between the various groups of participants, making it difficult to control for extraneous factors.

While some research questions can be answered with either of the two approaches, other hypotheses are better tested with one of these types of research only. In *intervention studies*, the method of choice is usually the longitudinal approach, so that the same group of participants is tested before and after an intervention takes place. Also known as *pre-test/post-test studies*, these studies can focus on a clinical, educational, behavioral, or cognitive intervention. For example, an intervention study with bilinguals may study the effect of language therapy on linguistic performance by having a bilingual child with language impairment take a battery of language tests before and after language therapy. The change in performance as a result of therapy is then examined. Another example is measuring test performance before and after enrollment in a dual-language immersion classroom. In both of these cases, taking a longitudinal approach and comparing performance of the same group before and after treatment is preferable to comparing performance of two different groups.

Finally, it is also possible to combine both approaches, if the research questions warrant doing so and if sufficient resources (time, participants, and money) are available. In the language therapy example above, the design can be altered from longitudinal to a combined longitudinal and cross-sectional design by testing two groups of bilingual children (with similar language impairments). One group receives language therapy in the first language and the other group receives language therapy in the second language. Pre- and post-intervention measures collected for both groups would allow cross-group comparisons that can elucidate (a) whether language therapy is effective for this particular language disorder in bilinguals, and (b) whether language therapy in one of the bilinguals' languages is more effective than language therapy in their other language.

Independent Variables, Dependent Variables, and Confounding Variables

In an experiment in which you study how a change in a certain variable influences performance, the variable that is being manipulated is called the *independent variable* and the variable that is being measured is called the *dependent variable*. For example, if you were interested in how language proficiency influences reading speed, you may want to design an experiment in which bilinguals with varying proficiency levels are asked to read text passages. In this case, language proficiency is the independent variable and reading speed is the dependent variable. The same variable can be either an independent variable or a dependent variable, depending on the design of the study. For example, in a study that focuses on the effect of age of acquisition of a second language on proficiency in that language, age of acquisition is the independent variable and language proficiency becomes the dependent variable. In short, the independent variable is always the one that is manipulated and the dependent variable is always the one that involves some type of measurement (e.g., score on a test, percent of words recalled, number of seconds it takes to complete a task, etc.).

The independent variable is usually varied across groups. That can be accomplished by either having different groups receive different conditions of the independent variable, or by having one group in which the independent variable is being manipulated (called the *experimental group*) and one group in which the independent variable is not being manipulated (called the *control group*). Experimental and control groups should be identical on all variables except the variable of interest, in order to ensure that whatever differences are observed between groups are genuine differences due to the independent variable and not due to other differences between groups or to placebo effects. *Placebo* effects (the term originates from medical studies that found that patients who were given a sugar pill, called a placebo, showed some clinical improvement in medical symptoms similar to those patients who received a real pill containing medication) in bilingualism research can arise from participants simply knowing that they are participating in a research study. Whenever possible, including control groups in your study is a good way to ensure its validity.

In the example considering language therapy for bilingual children with language impairment, performance on a language assessment scale is the dependent variable. The independent variable is language therapy. This independent variable could include multiple conditions, depending upon the design of the study. It could, for instance, have two conditions—treatment and no treatment—in which two groups of bilinguals are tested, one that receives language therapy and one that does not. Using a control group that does not undergo language therapy ensures that passage of time alone, without any treatment, is not responsible for improvements in performance. Or, language therapy could vary across three conditions—treatment in the first language, treatment in the second language, and no treatment—to compare the benefits of treatment in each of a bilingual's languages. Another condition that could be added to this study is a combination of first and second language use in treatment. In general, an independent variable can vary across multiple conditions, but, whenever possible, the most efficient and simple design that will answer the target question should be chosen.

In addition to independent and dependent variables, researchers are often faced with confounding variables. *Confounding variables* are variables that the experimenter did not plan to alter in the study design, but that nevertheless influenced participants' performance on the dependent variables *in addition* to the stated independent variables. Possible experimental confounds include participant characteristics, such as socio-economic status, gender, and language proficiency, as well as experimental variables, such as linguistic background of the experimenter, experimental setting, and stimuli selection. For example, participants may switch back-and-forth across languages more if the experimenter is bilingual than if the experimenter is monolingual; therefore a study that looks at code switches (overt verbal switches between a bilingual's two languages) should take into account the linguistic status of the experimenter. In short, a confound is a third variable that affects the outcome of the experiment.

At the same time, it is not possible to control for every single potentially confounding variable. When designing a study, consider the factors that are most likely to pose a problem for that particular research question and focus on those. A study is at greater risk for invalid and unreliable results if it does not take into account the relevant confounding variables in the design. Consider the example of studies reporting findings that bilinguals in the United States score lower than monolinguals on intelligence tests. Before you run forward with the conclusion that bilingualism is bad for you and that monolinguals are smarter than bilinguals, consider the fact that bilingualism studies from outside the U.S., such as the Quebec area of Canada (a bilingual French-English community) have failed to find differences in performance on IQ tests between bilinguals and monolinguals (Pearl & Lambert, 1962). Consider also that the studies reporting lower IQ scores for bilinguals did not take into account socio-economic factors such as family income and education. That, combined with the facts that the majority of bilinguals in the U.S. are Hispanic immigrants from Central and South America and that these groups are also frequently of lower socio-economic status, change the interpretation of that research entirely. Because socio-economic status is the confounding variable that drove the negative correlation, what it really tells us is that poverty (not bilingualism!) is bad for you and that linguistically diverse groups are disproportionately represented in the lower socio-economic brackets. Armed with a critical eye and a basic understanding of research, one can easily identify weak experimental designs, poor control over confounding variables, or unwarranted data interpretation. Generally speaking, the outcomes of a study (results and interpretation) are only as good as how the study was conducted (its methods).

Operational Definitions, Reliability, and Validity

For a variable to be valid and reliable, appropriate *operational definition* is necessary. *Operational definition* refers to the exact measure that is used to assess a particular construct. For example, if an experiment uses vocabulary size as its dependent variable, the operational definition has to indicate whether vocabulary size is defined as production vocabulary or comprehension vocabulary and what assessment tool or scale is used to measure it. One study may operationally define vocabulary size as a child's performance on the Peabody Picture Vocabulary Test (PPVT), a measure of comprehension. A different study may operationally define vocabulary size as all the words a child is producing, as reported by the parent. Of the two operational definitions, the latter would work better for a 13-*month* old, while the former would work better for a 13-*year* old. Note, however, that both measures assess the same variable—vocabulary size. Because different studies may use different operational definitions to measure the same dependent variable, it is important to pay careful attention to the operational definitions of both the dependent and the independent variables when designing or interpreting a study. In the case of vocabulary, for instance, one may specify that the variable of interest was vocabulary size, defined as comprehension vocabulary, operationally defined as performance on the PPVT, and measured by administering the PPVT in English by a licensed Speech-Language Pathologist.

Operational definitions are used to define constructs (such as vocabulary, or bilingualism, or creativity, etc.) in ways that are clearly measurable and that refer to observable behaviors, rather than abstract concepts. For example, one may want to study the effects of bilingualism on creativity. But what is creativity? A good operational definition and a way to reliably measure the behavior of choice are necessary in order for the study to be valid. If one defines creativity as the ability to maintain focus on a given task and then operationalizes it as the time one remains awake while reading a scientific paper, the creativity measurements obtained are invalid. The length of time it takes to doze off while reading a paper may be more indicative of the author's writing prowess, the reader's knowledge about and interest in the topic, as well as extraneous variables such as how much sleep the reader got the night before, how much coffee was consumed that day, and whether there will be a test on the material later, rendering this particular measurement of creativity both invalid and unreliable. A better way to test creativity may be to administer the verbal Torrance Test of Creative Thinking or one of the more recent tests designed to measure creativity.

A *valid* operational definition is one that measures precisely what it set out to measure. *Reliability* refers to the likelihood that the same finding will be obtained if the study is run again, either by the same or by a different researcher. If the construct of creativity AND the construct of bilingualism are carefully operationally defined, then any researcher who uses the same operational definitions, the same criteria, and the same tools, should be able to replicate the original experiment and obtain the same set of findings.

Note that in the example above, bilingualism is a construct that requires operational definition, as well. In fact, one of the most critical problems with bilingualism research is the lack of clarity in defining bilinguals and the lack of consistency in classifying different bilingual populations. Many individuals who are new to bilingualism research tend to group everyone who has any number of vocabulary words in another language as 'bilingual.' As a result, research results often appear contradictory when reporting experimental findings with bilinguals. The 'bilinguals' under study are sometimes foreign language learners who have never used their non-native language outside the classroom; other times these are fluent, equally balanced bilinguals who use both languages frequently in their everyday life; and yet other times they fall somewhere in-between, perhaps using both languages frequently, yet being more proficient in one than the other. And while attempts to define the different types of bilinguals by age of acquisition and language proficiency are not new (Ervin & Osgood, 1854, 1965; Weinreich, 1974), a consistent and universally-agreed upon classification of bilinguals is lacking in empirical reports. You may think that it is only a matter of labels, but consider the shortcut and clarity afforded to a researcher studying aphasia (a loss of language resulting from damage to the brain caused by injury or disease) by universally-agreed upon terminology to describe the aphasic population under study. Referring to participants as having "Broca's aphasia" or "Wernicke's aphasia" (for definitions of Broca's and Wernicke's aphasia, see Vaid's chapter on the neuropsychology of bilingualism in the current volume) makes many of the characteristics of the population in question evident, including affected areas of the brain and characteristic language deficits. Similarly, using a universal language to describe bilingual populations would increase the reliability and the validity of empirical studies. Until a consensus is reached on which labels to affix to bilingual groups that share certain characteristics, it is best to include any language history variables that describe the group under study when reporting a finding. This way, future replications of the findings are more likely since similar bilingual groups will be targeted for testing. In addition, by knowing what groups have already been tested, it becomes possible to extend a finding to other groups of bilinguals or second language learners that were not included in the population of the original study.

Between-Group, Within-Group, and Mixed Designs

Between-group (also called between-subject) designs include studies in which the independent variable varies across groups. Whenever more than one group of participants is tested and performance across groups is compared, the design of the study includes a between-group component. For example, whenever bilinguals are compared to monolinguals, or different groups of bilinguals are compared to each other, the design of the study is a between-group design. If there are only two groups tested (say, a bilingual experimental group and a monolingual control group), then the study is said to have one independent variable, "group," with two conditions (also called levels), experimental and control. If four groups are tested and compared to each other (say, an English-Japanese group with English as the native language, a Japanese-English group with Japanese as the native language, a monolingual English group, and a monolingual Japanese group) the study is said to follow a between-group design with an independent variable that has four conditions or levels.

Within-group (also called within-subject) designs include studies in which the independent variable varies within the same group of participants. Pre-test/post-test studies are one example of a within group design. In within-group studies, performance of a group of participants is compared to performance of the *same group* of participants under different conditions or at different points in time. For example, when the same group of bilinguals is tested in their first language and then tested again in their second language, the design of the study is said to be a within-group design. The number of times a measurement is made determines the number of levels a within-group independent variable has. So, if measurements are made twice, once in the first language and once in the second language, the study is said to follow a within-group design with an independent variable that has two levels. A study can have multiple independent variables at the same time. For example, in addition to first and second language, a study may include treatment status as another within-group independent variable. If a bilingual's performance is measured before, during, and after language therapy, for example, then the study is said to have a within-group design with an independent variable (i.e., treatment) that has three levels. A study that would combine both language (first or second) and treatment (before, during, and after) into the same design is said to have a within-group design with two independent variables. The first independent variable, language, has two levels and the second independent variable, treatment, has three levels. This is reported in scientific journals as a 2-by-3 design (multiply 2 X 3 and this would give you the number of possible levels), resulting in six conditions—(1) tested in the first language before treatment, (2) tested in the first language during treatment, (3) tested in the first language after treatment, (4) tested in the second language before treatment, (5) tested in the second language during treatment, and (6) tested in the second language after treatment.

Studies that incorporate both between-group variables and within-group variables are referred to as *mixed-design studies*. Mixed design studies include independent variables that vary both across the different groups tested, and within each group. For example, if the four groups mentioned earlier (an English-Japanese group with English as the native language, a Japanese-English group with Japanese as the native language, a monolingual English group, and a monolingual Japanese group) were tested each in different conditions (for example, before, during, and after a language intervention), the study would be described as a 4 x 3 mixed design, with two independent variables, where the first independent variable is group and is a between-subject variable with four levels, and the second independent variable is treatment and is a within-subject variable with three levels.

Any empirical study in the literature can be classified as between-, within-, or mixed-design. It is not unusual for individuals who are new to research to have difficulties identifying independent and dependent variables, their levels, and the study design. With enough practice reading research papers and designing experiments, this understanding quickly falls into place. Once internalized, the notions make it very easy to understand and process research. So that, for example, if you read a study that describes itself as a 2 x 2 mixed design with group (bilingual, monolingual) as a between-group variable and condition (picture prime, word prime) as a within-group variable, you can quickly use the learned heuristics to conjure up the specifics of that particular study. Thinking of an experiment in these terms aids both the researcher and the study's audience by facilitating clarity and precision.

Keep in mind that the more complex the design, the more difficult it becomes to control for extraneous confounding factors and interpret the results, making it challenging to conduct well-controlled empirical studies and limiting the reliability and validity of a study. Oftentimes carving a bigger question into its smaller components and designing simple, elegant experiments that test individual predictions is the optimal choice.

How Methodology Can Drive Outcomes in Bilingualism Research

The topics of study under the bilingualism umbrella are as diverse and the methodologies as numerous as those with monolinguals and include research in social science, biological science, and humanities. Around the world, scientists are studying bilingualism from developmental, cognitive, linguistic, neurological, psycholinguistic and sociolinguistic perspectives, and are focusing on topics such as cortical organization and neural processing of the two languages in bilinguals, acquisition of the two languages in bilingual children and second (and subsequent) language learning in adults, lexico-semantic representation and processing, language and memory, language loss/attrition and language interaction, communication disorders in bilinguals, etc.. In this section, we consider some of the most common difficulties and oversights that those new to bilingualism research face.

Representation and Processing of Languages in Bilinguals

The first thing to underscore when studying representation and processing of the two languages in bilinguals is that the terms *representation* and *processing* are not interchangeable, rather, they refer to different phenomena and should not be confused. Representation refers to the structure and organization of the different language components in bilinguals, while processing refers to activation of these components and their interaction within and between languages. Studies that focus on representation of the two languages usually focus on identifying, locating, and representing the organization of the two languages in the bilingual cognitive architecture. Studies of bilingual language processing usually focus on activation of the two languages (e.g., parallel or sequential), examining variables that influence the levels of activation of each language at each processing level, and the interactions within and between the two languages. For purposes of heuristics, it may be helpful to think of representation as a static phenomenon and of processing as an active one. The confusion between the two often lies in the fact that the same behavioral measures (e.g., reaction times) can be used to make inferences about both representation and processing. For example, how long it takes a bilingual to retrieve words that are spelled the same way, but carry different meanings in two languages (e.g., *pain*, which in French means *bread*) can be used to infer both how these words are stored in the bilingual lexicon, as well as how they are processed during access. If after seeing the word *pain* English-French bilinguals are faster to recognize the word *bread* in English, this can be interpreted as a result of overlapping representations for pain and bread at the orthographic and possibly lexical (but not semantic) levels. It can also be interpreted as a result of parallel processing in both languages (regardless of whether or not the representations overlap). In general, it takes a relatively sophisticated knowledge of psycholinguistics to be able to clearly delineate between the two when interpreting findings, and even then much of the thinking is inferential.

Second, when considering language representation and processing, it is important to distinguish between different levels of representation and processing. Higher levels of processing, such as mappings to the semantic level from the lexical level, may differ from lower levels of processing, such as mappings to the lexical level from orthographic/phonetic levels. Consider, for instance, Weinreich's model of lexico-semantic organization in bilinguals (see Figure 1, after Weinreich, 1974, pp. 9-10).

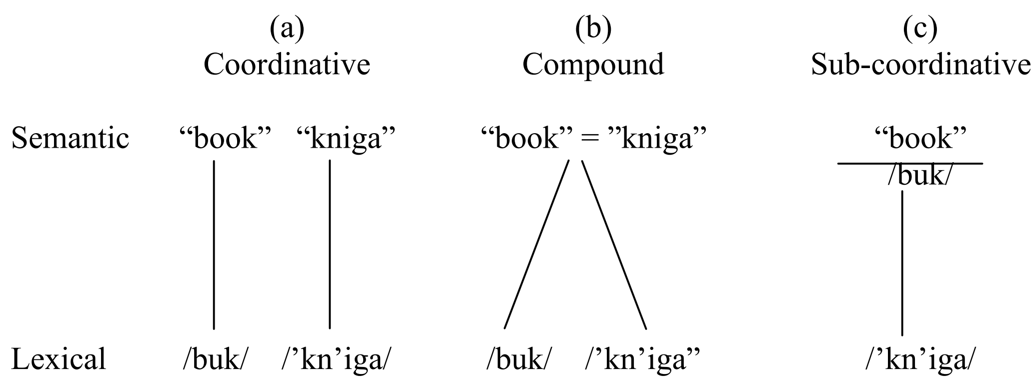

Figure 1. Weinreich's classification of bilinguals by mode of acquisition using Russian-English examples (after Weinreich, 1974, pp. 9-19).

Weinreich's discussion is written in terms of "signifieds" (conceptual structure/meaning at the semantic level) and "signifiers" (labels/words at the lexical level) and, according to the Weinrech model, bilinguals' lexico-semantic system can be organized in three ways, depending on how the languages were acquired. Coordinate bilinguals learn the two languages in separate environments and have one "signified" for every "signifier." Compound bilinguals learn their two languages in the same context concurrently and have only one set of "signifieds" with two signifiers for each signified. Sub-coordinative bilinguals interpret words of their weaker language through the words of their stronger language and the signifier in the first language becomes a signified for the second language. Multiple theories and models have been proposed to capture the nature of bilingual representation and processing, including such recent theories of language organization and processing in bilinguals as Kroll's (1993, 1997) Revised Hierarchical Model and de Groot's (1992) Distributed Feature Representation Model, both focused on higher levels of processing in bilinguals. On the contrary, models like the Bilingual Interaction Activation Model (Dijkstra, Van Heuven, and Grainger, 1998) and the Self-Organizing Model of Bilingual Processing (Li & Farkas, 2002) focus more on orthographic and phonological processing, respectively. Consequently, experimental paradigms that probe different levels of processing can lead to seemingly contradictory results. When designing or interpreting bilingual research, it is important to be aware of the levels of processing that the experiment taps into.

Third, bilingualism is not a static phenomenon, but a dynamic process that undergoes continuous change. As the level of proficiency and/or the manner of acquisition change, so may language processing and representation. Models that adjust representations based on new language experience capture the dynamic nature of bilingualism better than models where representations are pre-set and can not be changed. Moreover, it is possible that within a bilingual person, different representational systems coexist. Some words may be stored coordinatively, others—compoundly, yet others—subordinatively, depending upon the manner in which the words were acquired and/or upon specific characteristics of the words. Individual studies and general theories of bilingual language acquisition, representation, and processing must take into account the complex and fluid nature of the bilingual cognitive architecture when designing and interpreting studies.

Cortical Organization of Languages in Bilinguals

Another area of bilingualism research that frequently yields what appear to be contradictory findings is neurolinguistics. Neurolinguistic research with bilinguals usually focuses on hemispheric organization of the two languages and on location, size, and spread of activation associated with the two languages. The paradigms used include functional Magnetic Resonance Imaging (fMRI), Positron Emission Topography (PET), and ElectroEncephaloGrams (EEG), all of which rely on measuring cortical activity while brain resources are active. For instance, fMRI monitors the brain's use of oxygen from the blood, PET monitors the regional Cerebral Blood Flow and the brain's use of glucose from the blood, and EEGs measure electrical activity in the brain by detecting amount and size of event related potentials. The assumption driving these techniques is that the parts of the brain that use the most resources during a task are those that are most active and involved in performing that task, yielding a functional map of the brain. In addition, cognitive-behavioral studies with bilingual and multilingual individuals with aphasic language loss are used to assess changes in performance as a result of cortical lesion.

The topic that has received the most attention in neurolinguistic research with bilinguals is that of organization of the two languages in the brain, namely whether or not the same cortical areas subserve both languages or whether different areas are associated with each language. On the one hand, many studies have supported the position that the two languages are associated with different cortical structures. For example, selective disruption of first and second language naming has been reported in cortical stimulation studies (e.g., Ojemann & Whitaker, 1978). Selective language loss and/or differential language recovery in multilingual aphasic patients have also been interpreted as evidence for distinct cortical representations for each language (e.g., Gomez-Tortosa, Martin, Gaviria, Charbel, & Ausman, 1995; Nilipour & Ashayeri, 1989; Paradis, 1995; Paradis & Goldblum, 1989). On the other hand, a number of other studies have found evidence for overlapping cortical representations in bilinguals. Overlapping regions of activation in the left Inferior Frontal Gyrus were found in French-English bilinguals using PET (Klein, Milner, Zatorre, Meyer, & Evans, 1995). Chee, Tan, and Thiel (1999) found that late bilinguals activated the same regions for both languages in frontal, temporal and parietal lobes. And Illes et al. (1999) reported identical frontal lobe activations for both languages in Spanish-English bilinguals. Using functional neuroimaging in studies with bilinguals who acquired their second language post-puberty, Kim, Relkin, Lee, and Hirsch (1997) and Marian, Spivey, and Hirsch (2003) found that the two languages activated separate regions for each language within some cortical areas (e.g., Broca's area in the Inferior Frontal Gyrus, an area known to be particularly active during language production), but not others (e.g., Wernicke's area in the Superior Temporal Gyrus, an area known to be particularly active during language comprehension). This suggests that the two languages of a bilingual may or may not share cortical underpinnings, depending upon the specific task and processing involved, and that the networks of language activation in the brain are overlapping in some areas, but may diverge in others.

It becomes clear, then, that many of the differences in the neurolinguistic literature with bilinguals are due to differences in methodologies and/or participant populations. For example, studies with early bilinguals (bilinguals who acquired both languages in early childhood) have found mostly the same cortical areas activated for both the first and second languages (Chee et al., 1999; Illes et al., 1999; Kim et al., 1997; Perani et al., 1998). Studies with late bilinguals (bilinguals who acquired their second language later in life) produced mixed results, ranging from reports that the two languages are processed similarly (Chee et al., 1999; Illes et al., 1999; Perani et al., 1998) to reports that the two languages are processed differently (Dehaene et al., 1997; Kim et al., 1997). The variability is sometimes accounted for by taking into account language proficiency (Abutelabi, 2001). But even the distinctions in language proficiency and age of acquisition are rather global. In part, the differences in findings are also due to failure to distinguish between distinct activation patterns for different types of information, for example, orthographic, phonological, lexical, and semantic. Studies that focus on monolingual language processing have long been differentiating between different types and modalities of processing (Binder et al., 1994; Fujimaki et al., 1999; Nobre, Allison, & McCarthy 1994; Peterson, Fox, Snyder, & Rachle, 1990; Price & Giraud, 2001; Shaywitz et al., 1995), and similar distinctions should be considered when studying bilingual processing.

In sum, a global first-language versus second-language approach to the neurolinguistics of bilingualism– one that poses the general question of whether the same or different areas in the brain are associated with the first and second languages – is misleading. Bilingual language processing and the associated neural correlates are influenced by the type of processing involved, by the tasks and stimuli used, by experimental methodology, and by participant language history. Moreover, because there is not one area of the brain alone that is devoted to language processing, but rather a network of areas that work together, the question of same or different cortical activation in bilinguals is fundamentally misguided and needs to be reframed in more appropriate terms that reflect the network processing approach.

Language Development in Bilingual Children

Developmental research on bilingualism tends to generate broad public interest. From issues that affect bilingual children directly, such as educational placement of bilingual children, to issues that have the potential to influence the entire population, such as the requirement (or lack of it) of a foreign language in schools, research on bilingual children can impact educators, parents, funding agencies, and government policies. To illustrate the broad range of findings, consider the following two polar-opposite scenarios that happen every day in almost every school district in the United States. Every year thousands of middle-class and upper-class American children take a foreign language for purposes of enrichment. These children, their parents, and teachers function under the assumption that knowing another language "is good for you." At the same time and sometimes in the same school, thousands of other children, usually of lower-class and, sometimes, middle-class backgrounds, usually immigrant, often Hispanic or Asian, are discouraged from and sometimes forbidden to speak a native language, under the assumption that that will prevent them from mastering English and that, in general, raising children with more than one language "will confuse them" and will have long-lasting detrimental effects.

If pressed, both sides can provide what appears to be convincing evidence supporting their position. There are, for instance, studies on the impact of bilingualism on cognitive development that point out that bilingualism in children is associated with increased meta-cognitive skills and superior divergent thinking ability (a type of cognitive flexibility) and with better performance on some perceptual tasks (such as recognizing a perceptual object "embedded" in a visual background) and classification tasks (for reviews, see Bialistok, 2001; Cummins, 1976; Diaz, 1983, 1985). There are also studies that suggest that bilingualism has a negative impact on language development and is associated with delay in lexical acquisition (e.g., Pearson, Fernandez, & Oller, 1993; Umbel & Oller, 1995) and a smaller vocabulary than that of monolingual children (Verhallen & Schoonen, 1993; Vermeer, 1992). Both arguments are right in a sense, but before going on to discuss the methodological aspects behind such findings, lets digress here to provide assurance that, by all accounts, bilingual children catch up with their monolingual counterparts on tests of verbal ability by the time they are in middle school, and well-controlled studies provide no evidence for lower intellectual abilities of bilingual children compared to monolinguals. The early differences in linguistic performance of bilingual children can be attributed to a language development pattern that is somewhat different from that of a monolingual. Bilingual children learn earlier than their monolingual counterparts that objects and their names are not one and the same, that these are two separate entities and that one object can have more than one name. But whereas understanding that language is a symbolic reference system is advantageous for metacognitive development, it can be detrimental for early vocabulary development.

Consider, for instance, how language assessment usually takes place. If a monolingual child has three lexical labels for three semantic items, say 'milk', 'grandma', and 'dog,' and a bilingual child has two lexical labels in English (say 'milk" and 'grandma") and two in Spanish (say 'leche' and 'abuela', the Spanish words for 'milk' and 'grandma'), the monolingual child's vocabulary will be counted as consisting of three words and the bilingual child's vocabulary will be counted as consisting of two words. That is because vocabulary size is counted not as the number of lexical items known, but as the number of conceptual representations that have lexical labels. In this way, even though the bilingual child has four words in his/her vocabulary, the lexical labels map onto two conceptual representations only, compared to the three conceptual representations in the monolingual child.

This technique of assessment frequently places bilingual children at a disadvantage. Even worse, sometimes bilinguals are assessed in only one of their two languages, therefore providing an inaccurate assessment of the child's actual level of linguistic and cognitive development. For example, if the bilingual child in the example above knew four words in Spanish and four words in English, two of which in each language were translation equivalents, the child should be assessed as having a vocabulary of six words total (two concepts that had labels in Spanish only, two concepts that had labels in English only, and two concepts that had labels in both English and Spanish). Too often, however, children are assessed in only one of their languages, typically the language of the country in which they are being tested (e.g., English, often their second and less proficient language), therefore resulting in assessments that erroneously place the bilingual child at a lower level of cognitive development than his or her true level. The problem, although clearly disturbing, is not an easy one to fix. Most school districts and speech-language pathology clinics are physically and financially unable to test children in Polish, Romanian, Kirghis, Urdu, Quechua, or any number of languages that may serve as a child's native language. As a result, children may be subject to academic placement below their appropriate level, handicapping their later academic advancement (for more comprehensive discussions of first/second language knowledge and cognitive processing in bilingual children, see work by Cummins, e.g., Cummins, 1984, Chapter 2).

To this day, the majority of clinical and experimental studies focusing on cognitive development in bilingual children assess participants in one language only. However, when 'best performance' across the two languages is considered, thus assessing the highest level of cognitive development attained by a bilingual child (as opposed to the level reflected in one language), bilinguals do not appear to be at a disadvantage (e.g., Sheng, McGregor, & Marian, in press), and in fact may show some advantages. Measures of 'best performance' are measures of cognitive development that are not limited by the constraints of one language. For example, in a recent study comparing semantic organization in elementary-school age bilingual Mandarin-English children and monolingual English children, Sheng, et al. (in press) compared the number of syntagmatic and paradigmatic responses in a word association task. Paradigmatic responses (e.g., dog-pet, dog-cat, dog-collie) are indicative of a more mature semantic system organized around category relationships, whereas syntagmatic responses (e.g., dog-leash, dog-bark) are experience-based and characteristic of a less developed system (Nelson, 1985). Comparisons among children's performance in the first language, second language and best performances indicated that performance in one language, even when that was the dominant language, was not an accurate reflection of the child's level of semantic development, reinforcing the argument that assessment of bilingual individuals is most accurate when combining best performance across both languages.

Methodological Considerations

Now that the basics of research design and the necessary research vocabulary are in place, the next section will serve as a primer on what variables to consider when studying bilingualism, multilingualism, and second/foreign language learning.

Selection of Languages

Before deciding on the characteristics of the bilinguals to be tested in any given study, a more basic decision to make is what languages the bilingual participants should be speakers of. Aside from obvious practical considerations such as access to a subject pool and ability of the experimenter or research assistant to test in that language, another concern is proper choice of languages from a theoretical standpoint. The answer to some research questions may be directly influenced by the choice of languages that the bilingual group speaks. Languages of the world differ in their auditory qualities, written form, and cross-linguistic similarity. Not only do languages that are closer in a language family tree share origins, but they are also more likely to share vocabulary, phonological qualities, and alphabets. There are multiple language families (Eskimo-Aleut, Ural-Altaic, Central American, Hamito-Semitic, South American, Dravidian, Basque, Malayo-Polynesian, Sudanese-Guinean, Sino-Tibetan, Bantu, Japanese-Korean, Khoin, Austro-Asian, varying depending on grouping), each consisting of modern languages that share a similar language ancestor. Table 1 shows the proximity of some of the languages in the Indo-European Language Family. Bilingual speakers of two languages that are similar, such as Spanish and Portuguese, for example, or Dutch and Afrikaans, may yield dramatically different results in a study of spoken or written language processing, for instance, than speakers of two languages that belong to different branches of the same language family, such as German and Punjabi, or Welsh and Persian. The differences may be even more marked for languages from different language families, such as English and Mandarin, or Russian and Japanese.

Proto-Indo-European						
Germanic	Celtic	Italic	Balto-Slavonic	Indo-Iranian		Greek Armenian Albanian Tocharian Anatolian
		Latin		*Sanskrit*	*Iranian*	
German English Frisian Flemish Dutch Afrikaans Icelandic Norwegian Swedish Danish Faroese Gothic	Breton Manx Scots Gaelic Irish Gaelic Welsh Goidelic Cornish Brittonic Cumbrian Galatian Gaulish Celtiberian	Romanian French Portuguese Spanish Italian Catalan Occitan Sardinian Rhaetian	Russian Ukranian Belorussian Bulgarian Macedonian Seerbo-Croat Slovene Polish Czech Slovak Sorbian Latvian Lithuanian	Punjabi Lahnda Sindhi Pahari Dardic Gujarati Marathi Konkani Maldivian Sinhalese Hindi/Urdu Bihari Rajasthani Bengali Assamese Oriya	Ossetic Kurdish Persian Baluchi Tadzhik Pashto	

Table 1. The Indo-European Language Family tree, prepared using Baldi (1983) and Mallory (1991). Note that the components and organization of a language tree may vary depending upon criteria applied.

Moreover, some questions can only be answered with bilingual speakers of some, but not other, languages. If one were interested in studying how different alphabets influence processing of the same language, for instance, one would have to select a language that can be transcribed using more than one alphabet, such as Serbo-Croatian, which can rely on either the Latin or the Cyrillic alphabets. Some questions can only be answered by studying bilinguals whose two languages share alphabets (such as German-English bilinguals), other questions can only be answered by studying bilinguals whose two languages use different alphabets (such as Korean-English bilinguals), yet other questions can only be answered by studying bilinguals whose two written systems share some, but not all written symbols (such as Russian-English bilinguals). In studies of bilingual reading, for example, taking into account ease of grapheme-to-phoneme mapping is particularly important, because languages with shallow orthographies (consistent letter-to-sound mapping, like Spanish) are processed differently in some aspects than languages with deep orthographies (inconsistent letter-to-sound mapping, like English). When selecting the languages for a study, understanding the constraints that each language places on the research at hand and taking these constraints into account when interpreting findings is essential.

Selection of Participants

In general, research in psychology, communication sciences and disorders, education, and linguistics is not representative of individuals with a diverse linguistic and cultural background, despite the fact that, when the world as a whole is considered, bilingualism is the norm, rather than the exception (Grosjean, 1992; Harris & McGhee-Nelson, 1992; Romaine, 1995; de Groot & Kroll, 1997) and that the proportion of bilinguals in the United States is rapidly growing (Daw, 2002). It is estimated that there are about thirty times as many languages in the world as there are countries and that at least half of the world population is bilingual (Romaine, 1995). In the United States, the results of the 2000 Census indicate that, due to changes in ethnic, linguistic, and racial composition, the minority population (Spanish-speaking, in particular) is growing 12 times faster than the majority population and that the foreign-born population grew from 19.8 million to 30.5 million between 1990 and 2000. By the year 2000, 18% of American households spoke a language other than English at home, a proportion that is steadily increasing. And yet, the vast majority of experimental populations under study are monolingual. Linguistically diverse populations remain severely under-studied, under-served, and under-represented.

We have already mentioned that selection of target languages can influence the outcomes of a study and may lead to apparently contradictory results. But even in speakers of the same languages, a number of variables other than the language per se can influence research outcomes. For example, factors such as *age of acquisition* may influence cortical organization of a bilingual's two languages, with bilinguals who learned both languages in parallel from early in life showing more overlap in cortical areas associated with first and second language activation, compared to bilinguals who learned a second language later in life (e.g., Kim, Relkin, Lee, & Hirsch, 1997). *Manner of acquisition* is another factor that has been found to influence results of bilingualism studies, with bilinguals who learned their second language by rote memorization of translation equivalents in a classroom setting showing a lexico-semantic organization consistent with the Word Association Model (in which the meaning of a word in L2 is accessed via its translation equivalent in L1), while bilinguals who learned their second language via everyday use in a second-language-speaking environment showing a lexico-semantic organization consistent with the Concept Mediation Model (in which access to word meaning is by direct route from lexical label to concept in both the first and second languages) (Potter et al., 1984). Moreover, the actual conceptual representation of a word is more likely to differ in these two groups of bilinguals. Conceptual representations are more likely to be shared across languages for bilinguals who learned their second language in a classroom setting via translation of L1 lexical labels. Conceptual representations are less likely to overlap across languages for bilinguals who learned both languages via regular use in everyday environments (de Groot, 1992).

In addition to age of acquisition and manner of acquisition, another crucially important factor to consider is *proficiency*. How well do these bilinguals know their first and second languages? How well can they speak, understand, read, and write in their two languages? In addition to absolute proficiency in each of the two languages, what is the *relative* proficiency of one language with respect to the other—are these *balanced* bilinguals or is one language stronger than the other? In balanced and non-balanced bilinguals, relative activation of the two languages may differ, thus leading to different results in studies that compare activation and interference of the two languages in bilingual language processing. Relative activation of the two languages is likely to also be influenced by current use of and recent exposure to the two languages.

Psycholinguistic studies with bilinguals may require the researcher to also control for differences on such variables as working memory capacity, vocabulary size, verbal/non-verbal IQ scores, and/or other cognitive attributes as relevant to a particular experiment. Sociolinguistic studies with bilinguals may require the researcher to control for such differences as socio-economic status, level of education, gender, or birth order. When more than one group of participants is tested, control for potentially problematic confounding variables can be accomplished by matching participants in the two groups on relevant variables (e.g., non-verbal IQ scores) or by random assignment of participants to groups.

Selection of Tasks and Stimuli

There are a number of classic tasks in bilingualism research that have been shown to be valid and reliable measures of cognitive performance in bilingual studies. For example, picture naming, word translation, word recognition, passage reading, cross-linguistic priming, Stroop tasks, and more recently eye-tracking have all been used to study bilingual language representation and processing. Variations on classic monolingual tasks are frequently used. For example, priming tasks are frequently used with monolinguals to study lexical and semantic activation. A simple example would be presenting a participant with the prime "dog" and examining speed-of-recognition in a lexical decision task (is this a word or a non-word?) for targets such as "cat" and "cloud". Because "cat" is semantically related to "dog", it is usually recognized faster than "cloud," a semantically-unrelated item. In bilinguals, priming tasks are frequently used to establish whether semantic representations are shared between languages. For example, in a Spanish-English bilingual, would the prime "dog" activate the target "gato" (Spanish for "cat") in the same way that it activates the target "cat" and/or in the same way that the prime "perro"

(Spanish word for "dog") does? Similarities and differences in priming across languages and within languages provide insights into the extent to which the two semantic networks are integrated in a bilingual. Variations in the priming paradigm include presenting bilinguals with a prime that is either in the visual modality and can be a written word, a syllable, a letter, a symbol, a feature such as a line, etc., or in the auditory modality and can be a spoken word, syllable, phoneme, linguistic or non-linguistic sound such as tone or music. Similarly, the target can also vary within and across modalities, languages, and levels of processing that it taps into.

New tasks and methodologies are continually emerging. These, coupled with novel approaches or technologies, make it possible to study questions that were previously impossible to answer. For instance, in studies of bilingual spoken language processing, it has proved difficult to reliably measure activation of a non-target language without overtly using that language in a task. Recent developments in eye-tracking methodology make it possible to use bilinguals' eye movements as an index of language activation. For instance, showing that Russian-English bilinguals make eye movements to a *stamp* when instructed in English to pick up a *marker* (because the Russian word for *stamp* is 'marka'), suggests that bilinguals do not deactivate the non-target language during comprehension and thus supports a parallel-processing approach to bilingual spoken language comprehension (Marian & Spivey, 2003; Spivey & Marian, 1999). Activation of a language that is not overtly used is tested behaviorally by recording bilinguals' eye movements to objects whose names in the non-target language overlap with names of other objects in the target language. This use of eye-tracking technology to study non-target language activation in bilinguals provides the strongest support to date for parallel activation of both languages in bilingual spoken language processing. Of course, it is not the task's novelty that is important, but its ability to validly and reliably answer the questions at hand.

Virtually every task used to study monolinguals and virtually every approach in cognitive, behavioral, and neural sciences can be successfully applied with bilingual populations. Methodologies such as functional Magnetic Resonance Imaging, Positron Emission Tomography, or Evoked Response Potentials have all been successfully used with bilinguals. Challenges arise not because different tasks are used, but because the different tasks often probe different phenomena, and that is not always taken into account when interpreting the findings. For instance, a priming paradigm can be used to study phonological, lexical, or semantic processing and can show facilitation in one case and inhibition in another, underscoring the need to take into account the tasks that were used and the types of processing that were tapped into when obtaining a particular result.

In addition, careful consideration should be given to all stimulus characteristics that may bear on the outcome of the study. For example, most experiments with bilingual participants use linguistic stimuli, such as sentences, words, and phonemes. If the stimuli are words, for instance, factors such as word frequencies in the two languages may influence the outcome. Regardless of what languages or bilingual populations one uses, high-frequency words are likely to be recognized faster than low-frequency words, and this effect may influence the results. In addition to considering frequencies within a language, as is normally done in monolingual research, bilingualism research must also control for word frequencies *across* languages, to ensure that word frequency differences across languages are not driving the effects. Even when the actual stimuli used are not words, for example, when pictures are used, the label of the picture may continue to be important and factors such as word frequencies for the labels that the pictures represent should still be taken into account. Further, variables such as bigram frequency (the likelihood of two graphemes co-occurring together in a language) or consistency of letter-to-sound mappings in stimuli can also play a role, depending upon the question under investigation.

Running the Experiment

Before a researcher can begin running a study and collecting actual data, any research project with human participants must be approved by an Institutional Review Board (IRB) and all researchers must undergo training to comply with the Health Insurance Portability and Accountability Act. When testing participants, it is important to follow ethical guidelines and considerations, protecting participants' rights to privacy and confidentiality. Participants must provide informed consent and care should be taken so that no physical or psychological harm is inflicted upon them. When testing bilinguals, consideration should be given to which of the two languages should be used when obtaining informed consent.

Actual data collection can take place in many ways, including face-to-face format, computer-collected format, paper-and-pencil format, recordings of brain activity, etc. During experimental procedures, the guidelines set for language use throughout the experiment should be consistent with the purposes of the study. In some studies, language switches may be discouraged or even prohibited, while in other studies they may be acceptable or even encouraged. When analyzing data, one may work with videotaped or audiotaped recordings, written transcriptions, or digital files. In many cases, the raw data need to be transcribed and coded first. When transcribing and coding data, it is important to ensure that a coder is consistent within her/his own coding from time A to time B and that multiple coders are consistent across each other.

When testing a research hypothesis, most studies will rely on statistical analyses to ensure that their effects did not emerge by chance and are reliable (i.e., statistically significant). Statistical significance refers to the likelihood of a particular set of results to have happened by chance. For example, when reporting analyses that have a probability level less than 0.05 (reported as $p<0.05$), what the researcher is saying is that the likelihood of those results to have happened by chance is less than 5 in 100. In behavioral research, probability levels of 0.05 and 0.01 (one in a hundred) are typically used, while in medical research, more stringent criteria are usually used, with probability levels of 0.001 (1 in 1000) and 0.005 (5 in a 1000) more common. In addition or instead of significance testing, many of the more recent studies in the scientific literature have adopted the use of confidence intervals. A confidence interval assesses the probability that a parameter lies within a particular range of values. In other words, it provides a range of probabilities within which the true probability would lay 95% or 90% of the time, depending on the precision desired. It also provides a way of determining whether the sample is large enough to make the trial definitive. A narrow confidence interval implies high precision, making it possible to specify plausible values to within a tiny range. A wide interval implies poor precision, making it possible to only specify plausible values to a broad and uninformative range. For strong data analysis skills, there is no substitution for a good statistics course and good statistical analysis resources, such as software and textbooks.

Summary and Conclusions: The Journey to Scientific Paper

If you've ever stayed up all night reading a book because you couldn't wait to find out how it ended, or if you've ever been excited by the prospect of discovering or learning something new, your intellectual curiosity may make you a prime candidate for a research career, one marked by many hours at the library, in the lab, in front of the computer, and in the classroom. But there is a long road from that initial intellectual curiosity to a completed scientific project. And although the exact quest will differ in its content, with questions asked and specific studies varying widely, the path from the original idea to a shared scientific truth is very similar, at least for those working with bilinguals. It includes familiarizing oneself with the literature, defining a research question, considering relevant variables, running the experiment, analyzing data, and reporting the results. Once the data are collected and analyzed, the researcher approaches the end of the journey, when it's time to sit down and write up a report of the findings. You may think that the stack of papers reporting the outcomes of statistical analyses symbolizes the grand finale, but no research project is truly complete until its findings are written up and disseminated among the community of scholars working in that area. And whether it is a course term paper, a Doctoral dissertation, a conference paper, or a journal article, writing up the research project and transforming it into a scientific paper is for many the most challenging part of the research journey. Writing can be a difficult process; it requires discipline and organization. It is also the process that reveals any weaknesses of ideas and execution that may not have been apparent at the start of the project (a good reason to write research proposals early on). Many students have a difficult time at the writing stage because they expect perfection and are anxious about producing anything less than that, or because they feel overwhelmed by the scope of the task. Others will say that, once they know the results of the study, once they know 'how it all turns out,' they are no longer interested in the project. Whatever the reason, absence of a written report makes it difficult for findings to reach an audience, and without shared scientific knowledge, a project that has not been written up is about as good as one that has never been run.

If you find that the writing stage does not come as easily as you would have liked, try to come up with a few things that can help you accomplish your goals of a well-written, fine research paper. Figure out what works best for you as far as effective approaches to writing. Though the techniques that are successful vary from person to person, there are a couple of things that seem to work consistently. One is perseverance: Not giving up, staying on course, working on the project until it is completed. You may write more on some days, less on others, but be sure to write consistently. Make goals that are not overwhelming to start with. If it's going well, keep at it, if you are not particularly productive on a given day, work on the 'busy work', things that take time and effort but do not require as much thought. Another thing that many find helpful in early stages of a writing career is having someone to turn to for support. This may be a mentor, an adviser, a significant other, a parent, a student, a colleague, a friend, a former teacher, a therapist, a writing coach—anyone who is there for you when you need a word of encouragement and support. Eventually, as you become a more experienced writer and internalize the techniques that work for you and learn the ebb and flow of your particular writing style, you may no longer need a touchstone person to turn to, but in those early days, seeking out an environment that is supportive and encouraging may make the difference between a completed undergraduate honors thesis and graduating without distinction, or between a doctorate dissertation and an eternal ABD (all-but-dissertation). Finding out what and who motivates you to be a better writer and researcher can open the door to a successful research career for many years to come. And, as you benefit from the support and encouragement of others, remember to be that supportive and encouraging person to someone else in turn.

You will most likely go through numerous drafts before the final paper is published in a peer-review scientific journal. The first completed draft will be revised and polished many times before it is ready to be submitted to a journal or turned in to your advising committee. And the work does not end there. Members of advising committees often require revisions. Scientific journals follow a peer-review process and also, more often than not, require revisions. It is not unusual for an author to have to go through more than one round of revisions. Ultimately, the goal of that process is to produce an informative, well-written, high-impact study, expanding scientific horizons. Hopefully, this introductory overview of bilingual research methods can serve as a stepping-stone to understanding and conducting research on the human linguistic capacity and on the ability to accommodate more than one language simultaneously.

List of Key Words

Experimental studies, Descriptive studies, Correlational studies, Independent variable, Dependent variable, Confounding variable, Operational definition, Between-group design, Within-group design, Mixed Design, Longitudinal, Cross-Sectional, Reliability, Validity, Age of acquisition, Manner of acquisition, Proficiency, Early Bilinguals, Late Bilinguals, Balanced Bilinguals, Unbalanced bilinguals, Semantic, Lexical, Phonological, Orthographic.

Internet Sites Related to Bilingualism Research

http://www.communication.northwestern.edu/csd/faculty/Viorica_Marian/
http://www.edu.bham.ac.uk//bilingualism/database/biweb.htm
http://www.edu.bham.ac.uk//bilingualism/database/dbase.htm
http://www.ncela.gwu.edu/
http://www.lsadc.org/

Review Questions

For the experimental examples included in the observational and experimental studies section to the between- and within-group and mixed design section (2.1 to 2.5):

1. Identify the design, the independent variables and the dependent variables.
2. When conducting these studies, what confounds would you attempt to avoid and how?
3. How may the outcome of these studies vary between bilinguals who are highly proficient in both languages and bilinguals with different proficiency levels in the two languages?
4. Choose two of the examples used in the text and answer Review questions 1 to 3 for those sample experiments.

Suggested Research Projects

Experimental Study: The Stroop Effect with Bilinguals and Monolinguals.

Write color words using different color ink; for example write the word 'black' in either black or red ink. For half of the words the color of the ink and the color word should match, for the other half the color of the ink should be different from the color spelled by the word. Now ask your classmates to name the color of the ink for each word, ignoring the words themselves. Time the responses. Compare response times for items in which the color of the ink and the word match to items where the two do not match. How do you explain the results? Now ask those classmates who speak more than one language to name the color of the ink in the other language they know (not the language that the words are written in). Again, time their responses. Compare response times for trials in which the language used to name pictures is the same as the language of the written words to trials in which the language used to name pictures is different from the language of the written words. How do you explain these results? Did it make a difference whether the bilingual subjects were speaking their first or second language? Their stronger or weaker language? For more about the bilingual Stroop task, read Preston and Lambert (1969) and Tzelgov, Henik, & Leisner, (1990).

Observational Study: Bilingual Narrative Analysis.

Interview a bilingual speaker about a salient childhood event (e.g., a memorable birthday, a visit to the doctor, a move, a vacation trip, etc.) in each of the two languages s/he speaks (if you are not bilingual yourself, ask a bilingual classmate to conduct the interview). Tape the narratives. Does the bilingual switch to the other language during a narrative? Do the switches appear random or do they occur in a systematic fashion (e.g., at the same point in a sentence, or related to the same topic, or in one language more than in the other)? How do you explain these results? Are there other differences across the two languages (e.g., emotional content and intensity, self-construal)?

Suggested Readings

American Psychological Association. (2001). *Publication Manual of the American Psychological Association.* Fifth Edition. Washinton, D.C.: APA.

Baker, C., & Jones, S. P. (1998). *Encyclopedia of Bilingualism and Bilingual Education.* Clevedon, England: Multilingual Matters.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition, 1*, 131-149.

Ray, W. J. (`1993). *Methods: Towards a Science of Behavior and Experience.* Pacific Grove, CA: Brooks/Cole.

Weinreich, U. (1966). *Languages in Contact: Findings and Problems.* The Hauge-Paris: Mouton.

Author Notes

Preparation of this chapter was supported in part by grants NSF BCS-0418495 and NICHD 1RO3HD046952-01A1. The author thanks Margarita Kaushanskaya, Henrike Blumenfeld, Caitlin Fausey, and members of the Northwestern Bilingualism and Psycholinguistics Research Laboratory for feedback on earlier drafts. Correspondence about this chapter should be addressed

to Dr. Viorica Marian, Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, IL 60208-3570. Office phone: (847) 491-2420. Lab phone: (847) 467-2709. E-mail: v-marian@northwestern.edu

References

Baldi, P. (1983). *An Introduction to the Indo-European Languages.* Carbondale, IL: Southern Illinois University Press.

Ben-Zeev, S. (1977). The influence of bilingualism on cognitive strategy and cognitive development. *Child Development, 48*, 1009-1018.

Bialistok, E. (2001). *Bilingualism in Development: Language, Literacy, and Cognition.* New York: Cambridge University Press.

Binder, J.R., Rao, S.M., Hammeke, T.A., Yetkin, F.Z., Jesmanowicz, A., Bandertini, T.A., Wong, E.C., Estkowski, L.D., Goldstein, M.D., Haughton, V.M., & Hyde, J.S. (1994). Functional magnetic resonance imaging of human auditory cortex. *Annals of Neurology, 35*, 662-672.

Brigham, C.C. (1923). *A Study of American Intelligence.* Princeton, NJ: Princeton University Press.

Chee, M.W.L., Tan, E.W.L., & Thiel, T. (1999). Mandarin and English single word processing studied with functional magnetic resonance imaging. *Journal of Neuroscience, 19,* 3050-3056.

Cummins, J. (1976). The influence of bilingualism on cognitive growth: A synthesis of research findings and explanatory hypotheses. *Working Papers on Bilingualism, 9*, 1-43.

Cummins, J. (1984). *Bilingualism and Special Education: Issues in Assessment and Pedagogy*. San Diego, CA: College-Hill Press.

DeGroot, A. M. B. (1992). Bilingual lexical representation: A closer look at conceptual representations. In R. Frost & L. Katz (Eds.), *Orthography, Phonology, Morphology, and Meaning* (pp. 389-412). Amsterdam: Elsevier.

Diaz, R. M. (1983). The impact of bilingualism on cognitive development. In E. W. Gordon (Ed.), *Review of Research in Education,* Vol. *10*, pp. 23-54. Washington, DC: American Educational Research Association.

Diaz, R. M. (1985). Bilingual cognitive development: Addressing three gaps in current research. *Child Development, 56*, 1376-1388.

Dijkstra, A., van Heuven, W. J. B., Grainger, J. (1998). Simulating cross-language competition with the Bilingual Interactive Activation model. *Psychologica Belgica*, *38*, 177-196.

Ervin, S. M., & Osgood, C. E. (1954). Second language learning and bilingualism. *Journal of Abnormal and Social Psychology 49/Supplement*, 139-146.

Ervin, S. M., & Osgood, C. E. (1965). Second Language Learning and Bilingualism. In Charles Osgood & Thomas Sebeok (Eds.). *Psycholinguistics: A Survey of Theory and Research Problems With A Survey of Psycholinguistic Research*, pp. 139-146, Bloomington, IN: Indiana University Press.

Fujimaki, N., Miyauchi, S., Puetz, B., Sasaki, Y., Takino, R., Saakai, K., & Takada, T. (1999). Functional magnetic resonance imaging of neural activity related to orthographic, phonological and lexicosemantic judgements of visually presented characters and words. *Human Brain Mapping, 8*, 44-59.

Goddard, H.H. (1914). *Feeble-Mindedness. Its Causes and Consequences.* New York: Macmillan.

Gomez-Tortosa, E., Martin, E.M., Gaviria, M., Charbel, F., & Ausman, J.I. (1995). Selective deficit of one language in a bilingual patient following surgery in the left perisylvian area. *Brain and Language, 48*, 320-325.

Goodenough, F. (1926). Racial differences in the intelligence of school children. *Journal of Experimental Psychology, 9*, 388-397.

Grosjean, F. (1997). Processing mixed language: Issues, findings, and models. In A. DeGroot & J. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives*, pp. 225-254. Hillsdale, NJ: Lawrence Erlbaum.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition, 1*, 131-149.

Herrnstein, R.J. & Murray, C. (1994). *The Bell Curve.* New York: Simon & Schuster.

Illes, J., Francis, W.S., Desmond, J.E., Gabrieli, J.D.E., Glover, G.H., Poldrack, R., Lee, C.J., Wagner, A.D.. (1999). Convergent cortical representation of semantic processing in bilinguals. *Brain and Language, 70*, 347-363.

Kim, K. H. S., Relkin, N. R., Lee, K. M., & Hirsch, J. (1997). Distinct cortical areas associated with native and second languages. *Nature, 388*, 171-174.

Klein, D., Milner, B., Zatorre, R., Meyer, E., & Evans, A. (1995). The neural substrates underlying word generation: A bilingual functional-imaging study. *Proceedings of the National Academy of Science, USA, 92*, 2899-2903.

Kroll, J. F. (1993). Accessing conceptual representations for words in a second language. In Robert Schreuder and Bert Weltens (Eds.), *The Bilingual Lexicon* (pp. 53-82). Amsterdam: John Benjamins.

Kroll, J. F., & de Groot, A. M. B. (1997). Lexical and conceptual memory in the bilingual: Mapping form to meaning in two languages. In Annette M. B. de Groot and Judith F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives.* Mahwah, N. J.: Lawrence Erlbaum.

Li, P. & Farkas, I. (2002). A self-organized connectionist model of bilingual processing. In R. Heredia & J. Altarriba (eds.), *Bilingual sentence processing.* North Holland: Elsevier Science Publisher.

Mallory, J.P. (1991). *In Search of the Indo-Europeans: Language, Archaeology, and Myth.* New York: Thames and Hudson.

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition, 6,* 1-19.

Marian, V., & Spivey, M. (2003). Comparing Bilingual and Monolingual Processing of Competing Lexical Items. *Applied Psycholinguistics, 24,* 2, pp. 173-193.

Marian, V., Spivey, M., & Hirsch, J. (2003). Shared and Separate Systems in Bilingual Language Processing: Converging Evidence from Eyetracking and Brain Imaging. *Brain and Language, 86*, 70-82.

Nelson, K. (1977). The syntagmatic-paradigmatic shift revisited: A review of research and theory. *Psychological Bulletin, 84*, 93-116.

Nilipour, R., & Ashayeri, H. (1989). Alternating antagonism between two languages with successive recovery of a third in a trilingual aphasic patient. *Brain and Language, 36,* 23-48.

Nobre, A.C., Allison, T., & McCarthy, G. (1994). Word recognition in the human inferior temporal lobe. *Nature, 372*, 260-263.

Ojemann, G.A., & Whitaker, H.A. (1978). The bilingual brain. *Archives of Neurology, 35,* 409-412.

Paradis, M. (Ed.). (1995). *Aspects of Bilingual Aphasia*. Oxford, UK: Tarrytown.

Paradis, M., & Goldblum, M.C. (1989). Selective crossed aphasia in a trilingual aphasic patient followed by reciprocal antagonism. *Brain and Language, 36*, 62-75.

Pearl, E., & Lambert, W.E. (1962). Relation of bilingualism to intelligence. *Psychological Monographs, 76*, 1-23.

Pearson, B.Z., Fernandez, S.C., & Oller, D.K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning, 43*, 93-120.

Perani, D., Paulesu, E., Galles, N.S., Dupoux, E., Dehaene, S., Bettinardi, V., Cappa, S.F., Fazio, F., & Mehler, J. (1998). The bilingual brain: Proficiency and age of acquisition of the second language. *Brain, 121*, 1841-1852.

Petersen, S.E., Fox, P.T., Snyder, A.Z., Raichle, M.E. (1990). Activation of extrastriate and frontal cortical areas by visual words and word-like stimuli. *Science, 249,* 1041-104

Potter, M. C., So, K., Eckardt, V., & Feldman, L. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior, 23*, 23-38.

Preston, M., & Lambert, W. (1969). Interlingual interference in a bilingual version of the Stroop Color-Word Task. *Journal of Verbal Learning and Verbal Behavior, 8*, 295-301.

Price, C., & Giraud, A.L. (2001). The constraints functional neuroimaging places on classical models of auditory word processing. *Journal of Cognitive Neuroscience,13,* 754-765.

Romaine, S. (1995). *Bilingualism*, Oxford, U. K.: Blackwell.

Sarfarazi, M., Sedgwick, E.M. (1996). Event related potentials (ERP) as a function of language processing in bilinguals. *Electroencephalography Clinical Neurophysiology, 99,* 347.

Shaywitz, B.A., Pugh, K.R., Constable, R.T., Shaywitz, S.E., Bronen, R.T., Fulbright, R.K., Shankweiler, D.P., Katz, L., Fletcher, J.M., Skudlardki, P. (1995). Localization of semantic processing using functional magnetic resonance imaging. *Human Brain Mapping 2,* 149-158.

Sheng, L., McGregor, K., & Marian, V. (in press). Lexical-semantic organization in bilingual children: Evidence from a repeated word association task. *Journal of Speech, Language, and Hearing Research*.

Spivey, M., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science*, 10, 281-284.

Tzelgov, J., Henik, A., & Leisner, D. (1990). Controlling Stroop interference: Evidence from a bilingual task. *Journal of Experimental Psychology, 16,* 760-771.

Umbel, V.M., & Oller, D.K. (1995). Developmental changes in receptive vocabulary in Hispanic bilingual school children. In B. Harley (Ed.), *Lexical Issues in Language Learning* (pp. 59-80). Amsterdam: John Benjamins.

Verhallen, M. and R. Schoonen (1993). Lexical knowledge of monolingual and bilingual children. *Applied Linguistics 14*, 4, 344-363.

Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics, 22*, 217-234.

Weinreich, U. (1974). *Languages in Contact: Findings and Problems.* Paris: Mouton.