

Reported speech: Empirical corpus findings  
compared with EFL/ESL textbook presentations

by

Suzanne Emily Bayless Eckhardt

A thesis submitted to the graduate faculty  
in partial fulfillment of the requirements for the degree of  
MASTER OF ARTS

Major: English (Teaching English as a Second Language/Applied Linguistics)

Major Professor: Susan Conrad

Iowa State University

Ames, Iowa

2001

Copyright © Suzanne E.B. Eckhardt, 2001. All rights reserved

Graduate College  
Iowa State University

This is to certify that the Master's thesis of

Suzanne Emily Bayless Eckhardt

has met the thesis requirements of Iowa State University

\_\_\_\_\_

Signatures have been redacted for privacy

Signatures have been redacted for privacy

\_\_\_\_\_

## TABLE OF CONTENTS

ABSTRACT	vi
CHAPTER 1. INTRODUCTION	1
General Overview	1
Corpus Linguistics	3
Advantages of a Corpus-Based Approach to Linguistics	4
Prescriptive and descriptive grammar	4
Language intuition	5
Quantitative data	6
This Study's Role	7
Research Questions	8
Preview	9
CHAPTER 2. LITERATURE REVIEW	10
Reported Speech	10
Textbook Criticisms	11
Alternative Descriptions of Reported Speech: "New and Improved" Insights	13
This Present Study within the Context of Current Reported Speech Research	13
Corpus Linguistics	15
Intuition Alone or Combined with Large Amounts of Empirical Data?	15
Register Variation	17
This Present Study and Corpus Linguistics	17
The Big Picture: SLA Findings Related to Corpus Linguistics and Grammar	
Teaching: Implications for Reported Speech	18
Input and Communicative Instruction	19
How input relates to this study	21
Focus on Form	22
Focus on form defined	22
Focus on form evidence	22
How focus on form relates to this study	23
Consciousness Raising	24
Consciousness raising defined	24
How consciousness raising relates to this study	25
SLA Findings Related to This Present Study	25
Summary	26
CHAPTER 3. METHODOLOGY	27
Identifying Reported Speech	27
The Textbook Analysis	29
Data Collection	29
Description of Data	30
Data Analysis	31
Focus	31
Analyses	31
The Corpus Analysis	32

Data Collection	32
Selection of the corpus	32
Searching the corpus with computer programs	34
Hand editing of data	35
Data Analysis	38
Reporting verb frequency calculations	38
Analysis of verb tense combinations	39
Comparing the Textbook Findings and the Corpus Findings	40
Summary	40
 CHAPTER 4. RESULTS AND DISCUSSION	 41
Quantitative Findings	41
Textbook findings	41
Reporting verbs in textbooks	41
Verb tenses in textbooks	42
The embedded clausal verbs within the reported speech	42
The reporting verb's tense	45
Rules explicitly given in textbooks	46
Empirical Corpus Data Findings	48
Frequency of reported speech	49
Potential student problems with identifying reported speech	50
Verb occurrences	52
Register variation with reporting verbs	55
Tense combinations	56
Modals in the tense combinations	61
Register variations	61
Conditions influencing particular reported speech behaviors and patterns	64
Influential Present-Past conditions	65
Discussion	68
Answers to Research Questions	68
Research question #1	68
(1.a) Which tenses are used in reported speech?	68
(1.b) Which reporting verbs are used in reported speech?	69
(1.c) How do reporting verbs and tenses vary across registers?	70
Research question #2	71
(2.a) Which reporting verbs are used in textbook reported speech?	71
(2.b) Which tenses are used in textbook reported speech?	71
(2.c) Do these textbook presentations of reported speech reflect authentic usage patterns?	72
Reporting Verb Authenticity	72
Verb Tense Combination Authenticity	73
Authenticity of Reporting Verb Tense and Register	74
Register Variation Authenticity	74
Summary	74
Summary	75
 CHAPTER 5. CONCLUSIONS	 76
Key Points from This Study	76
Incorporating Both Textbook Information and Corpus Information	77

Implications of Research Questions and Findings	77
Textbooks vs. corpora	78
Implications for textbook presentation of verb tense combinations	78
Implications for textbook presentation of reporting verbs	79
Implications for textbook presentation of register variation	79
SLA-principled grammar teaching in light of findings	80
Reported speech grammar teaching and materials in general	81
Reported speech can be made less confusing	81
Textbooks should address register variation	82
Limitations	82
Future Research	83
Summary	84
 APPENDIX A. SAMPLE KWIC FILES	 85
 APPENDIX B. DETAILED VERB TENSE COMBINATIONS	 86
 APPENDIX C. RESULTS WITH “OTHER VERBS”	 87
 REFERENCES	 89
 ACKNOWLEDGEMENTS	 93

## ABSTRACT

This study uses the method of corpus linguistics to investigate a particular grammatical topic in English: Reported Speech. First, presentations of reported speech in seven ESL/EFL (English as a Second/Foreign Language) grammar textbooks are analyzed. The results of the analysis show what information ESL/EFL students currently receive from textbooks about reported speech. Second, a corpus study is implemented to determine how people actually use reported speech in real life. The results of the empirical corpus study represent authentic, naturally-occurring language and reported speech patterns in two registers of American English: Newspaper Writing and Conversation. Third, the results from the corpus study are compared with the results from the textbook analysis. This comparison establishes the extent to which the textbooks present real-life reported speech in terms of 1) reporting verbs used, 2) verb tense changes, and 3) register variation.

The findings reveal that some aspects of the textbooks' presentations reflect naturally-occurring language and other aspects do not. Most importantly, the empirical corpus findings point to register variation as an essential component in describing reported speech accurately. Reported speech behaves much differently in the Conversation register than in the News register. Reported speech also occurs more than three times as often in News than in Conversation. The textbooks do not address these register variations. Additionally, in all registers, the verb *say* emerges as the most common reporting verb, followed by *tell*. Most other reporting verbs presented in the textbooks are relatively infrequent. The verb tense "backshifting" process advocated in the textbooks occurs in only ~50 to ~70 percent of reported speech instances. Therefore, the alternative verb tense combinations, Past-Present, Present-Present, and Present-Past occur more frequently than the textbooks claim.

Findings from this study can be used to further inform our grammatical descriptions of reported speech. ESL/EFL textbook and materials writers can use these findings to prioritize their presentations of reported speech according to naturally-occurring usage patterns.

## CHAPTER 1. INTRODUCTION

### General Overview

This thesis uses the method of corpus linguistics to investigate a particular grammatical topic in English: Reported Speech. Reported speech represents a specific language function. In other words, we use reported speech to accomplish a specific language goal—to communicate a particular meaning. This reported speech language function occurs when a person reports what someone else has said. For example:

direct speech: *Linda: “I am hungry.”*

reported speech: *Linda said she was hungry.*

direct speech: *Bill: “I will pick you up at 7:30.”*

reported speech: *Bill told me that he’d pick me up at 7:30.*

It probably goes without saying that native English speakers use these sorts of sentences numerous times throughout their lives. Additionally, a proficient non-native speaker (NNS) (one who has learned English as a second or foreign language) needs to know how to communicate these reported speech ideas competently.

This grammatical topic of reported speech (a.k.a. indirect speech), however, has been identified as a complex, but often over-simplified area of English grammar. In particular, many linguists, teachers, and students have criticized how reported speech is presented to students who are trying to learn English as a second language (ESL) or English as a foreign language (EFL) (Goodell, 1987; Thompson, 1996; Yule, Mathis, & HopKins, 1992). The main thrust of the criticism is that the ESL/EFL grammar textbooks teach students to report speech in ways inconsistent with real life. In other words, native speakers often use reported speech differently from how ESL textbooks say they should. Critics claim that ESL textbooks either leave out important information, present false information, or both.

One of the main criticisms involves a “backshifting” of verb tenses into a past tense form. This refers to the fact that “*I am hungry*” becomes “*she said she was hungry*” in the above examples. Traditionally, textbooks claim that this backshifting does indeed happen;

textbook critics claim that this backshifting often does not happen, resulting in perfectly acceptable sentences such as “*she said she is hungry.*”

Another problematic issue with textbooks involves register variation. *Register* refers to a specific type of language used in a specific situation. Conversation, News, Academic Writing, Public Speaking are all different registers (Biber, Conrad, & Reppen, 1998). People use language in different ways depending on what situation, or register, they are operating in. A teacher most likely does not speak in front of his classroom in the same manner that he talks to his own parents on the telephone. A researcher does not use the same language to write academic papers as she does to talk to her neighbor. Recently, linguists have found such variation between and among registers to be a crucial aspect of accurate language description (Biber, 2001; Biber, Conrad, & Reppen, 1998). In fact, it may be impossible to describe language at all without using the concept of register variation. How can we even attempt to describe a “whole” language when language features behave so differently in most registers? Besides being an essential component of linguistic analysis, register variation is also one of the most fascinating aspects of language. Researchers have a plethora of language phenomena to study since the same language feature can behave quite distinctly depending on its situational context. Reported speech certainly is one of those language features that people apply in different ways depending on what register they are using. However, critics claim that textbooks do not mention register variation when discussing reported speech!

Although both the textbooks and their critics have valid points (points which will be addressed in further detail in the following chapters), it has been difficult for either side to support their claims because of lack of empirical evidence. Large amounts of naturally-occurring reported speech data have not been available. Therefore textbook authors and their critics have had to rely on their own beliefs about reported speech behavior. These beliefs have been based on the reported speech occurrences that researchers happen to encounter in real life or think up in their own minds. Perhaps these beliefs provide a good starting point for investigating reported speech. However, a truly reliable description of reported speech needs to be based on large amounts of representative, empirical data. The data should be extensive enough to represent various registers and principled enough to generalize findings. Until recently, obtaining such quality data has been difficult. Even if such data were obtained,



sorting through the data had been an even more challenging task. But advances in computer technology have made it possible to gather and analyze these large amounts of quality data. This method, called corpus linguistics, becomes useful at this point in our discussion of reported speech.

## Corpus Linguistics

Corpus Linguistics refers to a method of analyzing language using a corpus (or corpora, if more than one corpus). A linguistic corpus is a large collection of language that has been stored on a computer to be accessed by a linguist. Linguists build corpora by collecting thousands of naturally-occurring language samples. It is very important that the language samples they gather come from real-life language contexts. The corpus data should be a collection of *naturally occurring* language—language that people have actually written or spoken for real communication purposes. A corpus can include language from any real-life situation where language is used: casual conversation, formal speeches, service encounters at a store between a clerk and customer, spoken news reports on the radio or television, written news reports in magazines or newspapers, academic writing, fiction from novels, letters to relatives, etc. Once a reliable, representative corpus has been formed, linguists can study numerous language topics in a more reliable way than was previously possible. Linguists are able to search for (or program a computer to search for) a particular language feature and then find thousands of samples of that feature in real, authentic contexts.

In order to trust the results, however, linguists must ensure that the corpus is reliable. A reliable corpus means that the corpus is representative; it draws from numerous geographic/dialectal locations and encompasses various registers. A good corpus also keeps track of all these variables. Commonly, a reliable corpus will distinguish between registers such as News, Conversation, Academic Prose, or other register categories. A reliable, representative corpus could also make it possible for a researcher to distinguish personal attributes—for example conversation among men vs. conversation among women—and compare various language features between the two registers. So a reliable corpus that

represents different language varieties becomes quite a valuable, powerful tool for language analysis.

Although unquestionably an invaluable tool for linguistic research, no specific corpus can claim to represent a total language. A corpus is only as representative as its contents. Findings from a corpus of academic writing cannot be generalized to other types of discourse such as fiction or political speeches. And findings from a corpus of written News are not necessarily representative of all written News language unless the corpus contents are truly representative. For instance, a corpus that consists of editorial pieces from Midwestern U.S. newspapers cannot be said to represent all of written News. To represent all of written News, the corpus would need to include multiple sections of diverse newspapers from many regions. However, that corpus of Midwestern editorial pieces might be representative of its specific genre: Midwestern editorial writing. So if a corpus has been built in a principled manner, then one can assume that the corpus data actually does represent the specific type of language in the corpus. Researchers should be extremely careful to choose a representative corpus; not all existing corpora are suitable for all research endeavors. Yet researchers should be careful to avoid overgeneralizing corpus results to language varieties not represented in the corpus contents.

### **Advantages of a Corpus-Based Approach to Linguistics**

I use corpus linguistics in this study because a corpus-based approach to studying language use has been shown to have advantages over traditional approaches:

#### **Prescriptive and descriptive grammar**

One traditional approach exists as a *prescriptive* type of grammar. Prescriptive grammar claims that certain language forms are superior to others—for example “is not”/“am not” is superior to (or more “standard” than) “ain’t;” or “she said she was hungry” is superior to “she says she was hungry.” Prescriptive grammar makes judgements about language; it purports rules governing “proper” language use such as “never end a sentence with a preposition.” It *prescribes* how to use language “correctly” and aims to perfect a “standard” sort of language. In contrast, a *descriptive* type of grammar does not aim to judge the inherent

goodness of a language form. Rather, descriptive grammar objectively reports actual language usage regardless of whether or not prescriptionists approve of the “correctness” of the language. A descriptive grammar acknowledges all versions of a language feature that occur whether they are considered standard or nonstandard and goes on to describe scenarios, situations, and reactions to all versions. A prescriptive grammar, however, only shows interest in the one correct version. Yet both prescriptive and descriptive grammars have value. Even though it may be random—or even absurd—that prescriptionists have judged one language feature as superior, one must acknowledge that some awareness of prescriptive grammar rules is necessary for social pragmatic reasons. After all, using “non-standard” language forms can have very real consequences and social ramifications.

Within this basic distinction between prescriptive and descriptive grammar, corpus linguistics falls under a descriptive type of linguistic approach. Having an inherently descriptive nature, a corpus-based approach is more objective than traditional prescriptive approaches. This results in a more accurate language description of real language since a corpus-based approach accepts all types of language use as worthy of being studied. Corpus linguistics (or any descriptive linguistic approach) does not disregard “nonstandard” language forms, but instead seeks to analyze and describe those forms as interesting phenomena. Therefore, one major advantage of a corpus-based approach lies with its descriptive nature as opposed to a more limiting prescriptive traditional approach.

### **Language intuition**

Another advantage of a corpus-based approach involves the idea of language intuition. When using some traditional approaches, linguists must rely on their own memories (or the memory abilities of their colleagues or informants) to inform them of how a language feature is used. Linguists using this approach trust instinct—either their own or others’ “native speaker instincts.” This notion holds that any native speaker of any language inherently knows how the language can be used; it relies on native speakers’ brains subconsciously internalizing the underlying structure of the language. Relying on language *intuition*, as this would be referred to, can be quite useful, but also potentially dangerous since our intuitions alone are not accurate with some details. We do not notice every aspect of language and are incapable of

remembering everything that we do notice. We tend to remember unusual language uses and so we often overlook the common, ordinary language patterns that continually surround us. So while we may be able to trust native-speaker-intuition to tell us which word fits more appropriately in a certain phrase, we cannot rely on that intuition alone to identify the most commonly used word in spoken and written English.

Relying on our own opinions of language use also runs the risk of reflecting our inappropriate, or irrelevant, judgments about the “correctness” of language. So even though people might say “*I don’t got none*” more often than they say “*I don’t have any*,” our preconceived ivory-tower notions about the “incorrectness” of the first phrase may influence our descriptions of the language feature and distort the language reality. This reflects the previously discussed debate between *prescriptive* and *descriptive* grammar. Intuition can be used in both prescriptive and descriptive ways.

Despite its limitations, intuition can provide linguists with valuable information; intuition has shown to be a remarkably useful tool. Yet intuition becomes even more powerful when used in conjunction with a corpus-based linguistic approach. Our intuitions often provide the linguistic questions to investigate using a corpus-based approach. These intuitions can also guide the research process throughout a corpus study. A combination of intuition and corpus linguistics is more reliable than either approach alone. A corpus-based approach can deal with the linguistic questions that are beyond the scope of human intuition, and intuition leads corpus studies to explore relevant topics. Intuition often provides the impetus for the most interesting corpus studies. Corpus-based techniques allow us to check out our intuitions as well as answer otherwise impossible linguistic questions. Therefore, the second major advantage of corpus linguistics is its ability to work with and enhance language intuition, but also to account for intuition’s weaknesses.

### **Quantitative data**

A third advantage that a corpus-based approach has over traditional approaches to language study lies in sheer numbers. Even when more traditional approaches incorporate empirical, naturally-occurring examples, it is difficult for a person to keep track of the numerous occurrences and factors that a corpus-based approach sorts out. Exactly how

common is the feature? How often does it occur? What does it occur with? Does its use vary between different varieties of language? With a corpus-based approach a researcher's computer can categorize in minutes what would have taken the researcher days to do by hand. With such incredibly fast access to millions of words, corpus linguistics enables us to quantify language, thereby addressing formerly-impossible questions with much more confidence.

## **This Study's Role**

Since this study uses a corpus-based approach to find out more about reported speech, its results, by the very nature of corpus linguistics, describe naturally-occurring reported speech. Whatever results surface will represent “real language usage”—at least real language within the context of the corpus used.

The study consists of two parts: 1) an analysis of naturally-occurring reported speech in a large corpus, and 2) an analysis of the presentation of reported speech in current ESL/EFL grammar textbooks. The two analyses are compared to see to what extent the textbooks' descriptions represent actual language use. In other words, the textbooks' intuitively-based prescriptions about reported speech will be checked with the descriptions of reported speech obtained from the corpus's data. We will then be able to test, in some sense, whether the criticism of ESL textbooks is justified; the resulting data should enable us to see if textbooks really are off-base or not. In particular, the study focuses on:

- the frequency of different reporting verbs
- the sequences of verb tenses in reported speech
- variation across two varieties of language: spoken conversation and written news

Findings from this study should also apply to ESL grammar teaching. Effective grammar teaching should be based on principles from second language acquisition (SLA). Successful grammar teaching utilizes methods which help students acquire a second language. SLA research shows that students benefit from abundant exposure to accurate, authentic language (i.e., “input”). A corpus-based description of reported speech can provide students with such exposure. SLA research also shows that students benefit from analyzing that

accurate, authentic language in certain ways—through particular teaching methods such as Focus on Form and Consciousness Raising. A corpus study of reported speech can provide ideal data for teachers to implement focus on form and consciousness raising. So, implicit in the rationale for this study is the idea that ESL grammar materials need to be consistent with sound teaching principles—principles that are grounded in SLA theory. Because of this assumption and its implications, we will delve into a sampling of SLA grammar teaching research in the literature review section (chapter two).

## **Research Questions**

The main reported speech issues that textbooks are criticized for guide this corpus investigation. Textbooks take criticism primarily because of three issues: reporting-verb choice, verb tense combination, and register variation. Therefore, this study addresses the following research questions:

### **RESEARCH QUESTIONS:**

- 1) What does an empirical corpus study reveal about the usage patterns of reported speech in American English?
  - a. Which verb tense combinations are used with naturally-occurring reported speech? (i.e., Does backshifting occur?)
  - b. Which reporting verbs are used most commonly in naturally-occurring reported speech?
  - c. How do reporting verbs and tenses vary across two registers: Conversation and Newspapers?
- 2) How do EFL/ESL textbooks present reported speech?
  - a. Which verb tense combinations are used in textbook reported speech? (i.e., How is backshifting presented?)
  - b. Which reporting verbs are used in reported speech?
  - c. Do these textbook presentations of reported speech reflect usage patterns found in naturally occurring speech?

## **Preview**

The next chapter provides a fairly extensive literature review that synthesizes relevant research findings about important topics in this study—reported speech, corpus linguistics, SLA with grammar teaching. Chapter three explains the design and methodology of the study. Chapter four presents the empirical results and discusses their interpretations, and chapter five synthesizes conclusions and proposes implications.

## CHAPTER 2. LITERATURE REVIEW

This chapter gives an overview of other research relevant to my topic. The literature review integrates the topics of reported speech, corpus linguistics, and second language acquisition findings related to grammar teaching and ESL materials. I then position my current research study within a framework of the research that has been done so far.

### Reported Speech

Reported speech, as described in chapter one, has been a subject of controversy among grammarians. Educators, materials writers, and linguists have recently questioned traditional grammatical prescriptions about reported speech. Though these technicalities of reported speech will be dealt with in more detail in later chapters, most grammars, traditionally, have advocated a backshifting of tenses (the main clause's tense and the embedded clause's tense) in reported speech. To be more exact, this means that if a person heard someone make a statement in present tense, then the person needs to "report" the speech in a past tense form. "I am hungry" would be reported as "She said she was hungry."

While the literature on reported speech could not be described as "abundant," certain researchers have found this topic worth addressing. Below I review two broad categories of articles: 1) criticism toward current reported speech descriptions and 2) suggestions for improving reported speech descriptions. The first category of research criticizes the way in which reported speech is described in general and presented to students; these researchers claim that both the descriptions and presentations are inaccurate (Goodell, 1987; Yule, Mathis, & HopKins, 1992). The other broad research category proposes alternative, "new and improved" descriptions of reported speech; these researchers claim that their proposals are more accurate than the traditional descriptions (Comrie, 1986; Declerck & Tanaka, 1996;



Huddleston, 1989; Jacobs, 1999; Yule & Mathis, 1992). Often these alternative descriptions involve some naturally occurring language samples as evidence. This added element of authenticity enables the researchers to have more confidence that their findings are closer to real language use than textbook information. However, the studies do not approach the magnitude of a large-scale corpus study.

Another strand of related research involves researchers who attempt to describe reported speech and reported-speech-type forms less judgmentally—merely as interesting phenomena (Blyth, Recktenwald, & Wang, 1990; Buttny & Williams, 2000; Ferrara & Bell, 1995; Jacobs, 1999; Johnstone, 1987; Rings, 1992; Romaine & Lange, 1991; Tagliamonte & Hudson, 1998; Thompson, 1996). These researchers incorporate a wide range of topics including racism, authoritarian attitudes displayed in “say” or “said”, news reporter agreement, and the recent popularity of “I’m like” and “s/he goes” to report speech or recreate dialogue among younger generations. While this strand of research is quite interesting, it does not directly relate to primary issues of this present study, and so it is not outlined any further in this literature review.

### **Textbook Criticisms**

Most directly relevant to this study are the assessments and criticisms of reported speech presentations to students. Goodell (1987) describes ESL textbook reported speech as “incomplete,” and she proposes “a more explicit description” that encourages students to make choices about reported speech tenses based on more global, semantic practical considerations of the situation instead of unquestionably applying the Sequence of Tenses backshifting rule (p. 305). She asserts that students struggle with reported speech, at least in part, because it is presented to them inaccurately (or incompletely) (p. 306). Goodell gives several examples where applying the Sequence of Tenses backshifting process in the reported speech would distort the original utterance’s meaning. She ultimately proposes that grammatical components of reported speech must change to appropriately reflect a proper “spatiotemporal” point of view for the reporter (p. 312). Goodell outlines one overarching rule to encompass potential reported speech changes:

In indirect speech, elements such as pronouns, verbs, and adverbs from the original

utterance must conform to the here and now of the act of reporting. Whether or not this necessitates a change from the original utterance depends on the relevance of the pointing (deictic) qualities of these words for the present act of reporting (p. 313).

In short, Goodell advocates a more flexible, more practical presentation of reported speech that allows students to understand the reasons behind direct-indirect speech changes. This enables students to reflect on and consider the impact of their reported-speech grammatical choices instead of blindly following a mechanical backshifting rule or other rules.

The importance of this [Goodell's] rule lies in the fact that it emphasizes the underlying theoretical principles behind the adjustments in the indirect reported speech. This analysis provides a coherent explanation for exceptions to the rules in standard grammars, and it enables the student to develop an understanding of the underlying semantic motivation of reported speech in English (p. 313).

Such a thoughtful alternative displays a refreshing open-mindedness about the varied potential manifestations of language functions. Yet within Goodell's enlightening article, only one potential criticism exists. Goodell mentions that, often, "the goal of the report is to recreate the atmosphere originally present," but sometimes remarks that using the verbs *say*, *tell*, and *ask* result in a "flat" depiction of the event (p. 316, p. 315). While her assessment may be true, it also seems to be a problematic, controversial insertion of personal preference or language style. Goodell seems to imply that other "non say-tell verbs" are stylistically superior. It is true that other nongeneric verbs may be more specific for reporting speech than *say*, *tell*, and *ask*, but it is also dangerous to imply that this quality should make the nongeneric verbs common, especially in spoken, conversational contexts. It seems that *say*, *tell*, and *ask*, though generic, are probably much more frequent than their more specific, nongeneric counterparts if only because the nongeneric verbs seem peculiar—too formal—for most conversational functions. My present research study specifically addresses this question of reporting verb frequency.

Yule, Mathis, & HopKins (1992) also criticize textbooks' limited analyses of reported speech, claiming that many non-textbook grammatical forms are used in reality. They claim that many textbook explanations "are quite inadequate accounts of what [students] are likely to encounter in...spoken and written English outside the classroom" and suggest that students be

exposed to a wider “range of options” for reporting speech (p. 245). They go so far as to accuse customary textbook instructions, which present reported speech as stemming directly from direct speech, as being “a dangerous illusion” and point out that ESL textbooks fail to show “a reporting format which quite naturally mixes features associated with both direct and indirect speech” (p. 246). In other words, their recurring, basic point expresses a familiar, resounding message about reported speech: it’s not so simple. ESL textbooks present an oversimplified version picture of reported speech that just doesn’t hold true in real life language.

### **Alternative Descriptions of Reported Speech: “New and Improved” Insights**

By studying the phenomenon of this language function, many researchers have given us valuable insights into the numerous manifestations of reported speech grammatical form. While some insights have been rather obscure, others have been major enough to be repeatedly demonstrated in other literature. Table 1.1 summarizes many of these findings (see next page).

In sum, all of these studies recognize our lack of knowledge about authentic reported speech. Comrie (1986), Declerck & Tanaka (1991), Huddleston (1991), and Jacobs (1999) argue about (both for and against) the backshifting rule specifically. Buttney & Williams (2000), Rings (1992), Thompson (1996), and Yule & Mathis (1992) remind us of the wide variety of ways that reported speech occurs. They encourage us to broaden our perspectives concerning reported speech characterizations.

### **This Present Study within Context of Current Reported Speech Research**

The research outlined above attests to the complexity of this reported speech language function. The research also points to the inadequacy of current descriptions about reported speech. While the research studies to date that have addressed reported speech have given many valuable insights, there is still a glaring need for an extensive, quantitatively reliable research endeavor to investigate the behavior of reported speech. Most of the previous studies have used either artificially constructed hypothetical examples of reported speech or have used naturally-occurring data samples from random reported speech occurrences happened upon or from small corpora. What we need are large-scale corpus studies into this matter. In order to

Table 1.1 Various Research Findings about Reported Speech Behavior

RESEARCH STUDY	FINDINGS
Declerck & Tanaka (1991)	<ul style="list-style-type: none"> <li>- Certain factors prevent using a present tense clausal verb in reported speech.</li> <li>- This "casts doubt on the traditional Sequence of Tense" rules (p. 283).</li> <li>- Also, a past tense form (a relative tense) is unmarked and a present tense form (an absolute tense) is marked.</li> </ul>
Jacobs (1999)	<ul style="list-style-type: none"> <li>- Uses a small corpus of press releases (from <i>Exxon Corp.</i>)</li> <li>- Findings confirm both conditions of Declerck &amp; Tanaka's analysis:</li> <li>- present tense is only used in the complement clause when 1) there is no past temporal focus and 2) the reporter has some positive reason for representing the proposition in the complement clause as <u>true</u> in the other's (own) intensional domain</li> <li>- Using the marked present tense seems to indicate agreement (between reporter and the statement)</li> <li>- Past tense is used whenever there is a past temporal focus and with a reporter's disagreement</li> <li>- "Immediacy of interest" also influences present tense usage in News</li> <li>- Findings also discuss when the r.s. is about the reporter and also issues of giving authority in public relations.</li> </ul>
Yule & Mathis (1992)	<ul style="list-style-type: none"> <li>- Demonstrates two types of "hybrid-alternative" speaker reporting methods involving "staging" and "constructed dialogue"</li> </ul>
Comrie (1986)	<ul style="list-style-type: none"> <li>- Shows that reported speech tenses are determined by the Rule of Sequence of Tenses</li> <li>- Argues that a clausal verb is in the past tense in reported speech because it follows a past tense reporting verb, NOT because it has a past time reference</li> </ul>
Huddleston (1989)	<ul style="list-style-type: none"> <li>- Examines both Comrie (1986) and Declerck (1990), claiming that both "omit certain crucial types of examples"</li> <li>- Asserts that "we need some combination of a semantic and syntactic account" (i.e., considerations for both a backshifting rule and Declerck's semantic reasoning) (p. 335)</li> </ul>
Buttny & Williams (2000)	<ul style="list-style-type: none"> <li>- Reveals that participants in a discussion about racism use reported speech in particular ways to recreate the power of someone else's words</li> </ul>
Thompson (1996)	<ul style="list-style-type: none"> <li>- Gives a broader range of types of language reports --- one of which is classic reported speech;</li> <li>- it "serves as a...reminder of how much more variation there is in reporting, even in structural terms, than allowed for in the traditional view" (p. 523)</li> </ul>

obtain a representative sample of authentic reported speech, a large-scale, diverse, purposeful, principled, professional corpus should be used.

This present study fits in at this point to fill this gap. Although my proposed study will not and cannot answer all the questions about the multifaceted components of reported speech, it does provide a systematic empirical exploration into reported speech. The study uses the previous reported speech literature as a foundation and as a context to contribute within. It presents a relatively large-scale study of reporting verb choice and verb tense combinations using corpus linguistics. Literature concerning corpus linguistics is therefore discussed in the next section.

## **Corpus Linguistics**

### **Intuition Alone or Combined with Large Amounts of Empirical Data?**

As mentioned in the introduction, corpus linguistics involves using a computer to store and sort large amounts of language for linguistic analysis. The large amounts of language encompass millions of words and should come from a representative variety of sources and account for a satisfactory range of geographic areas, speakers, dialects, registers, etc. A good corpus contains language from (and can distinguish among) different registers such as: spoken news; written news; academic prose; fiction; conversation including service encounters at stores, dialogues between friends, relatives, and strangers—conversation in formal and informal settings, with a range of conversational topics, etc. A key component which distinguishes corpus linguistics is its ability to provide researchers with huge amounts of representative, naturally-occurring language. This sort of corpus data has an advantage over the more traditional, limited scope of fabricated examples that grammarians conjured up from their minds and own language experiences. Sinclair (1991) describes this former, traditional situation as similar to the state of the physical sciences 250 years ago:

The tradition of [previous] linguistics [had] been limited to what a single individual



could experience and remember ... Starved of adequate data, linguistics languished—indeed it became almost totally introverted. It became fashionable to look inwards to the mind rather than outwards to society. Intuition was the key, and the similarity of language structure to various formal models was emphasized. The communicative role of language was hardly referred to (p. 1).

Jan Aarts (in Aijmer & Altenberg, 1991) characterizes these different types of approaches as resulting in “intuition-based” and “observation-based” grammars (p. 44). Of these two types of grammars, corpus data is more realistic and unbiased, and the sheer amount of data makes corpus-based linguistic analyses more reliable than most traditional linguistic studies alone. Biber, Conrad, & Reppen (1998) express this sentiment by showing that “empirical investigations of corpora can shed new light on previously intractable research questions in linguistics” (p. ix).

Basically, corpus-based studies have been able to enhance conventional linguistic studies in that “the corpus-based approach ... provides a means of handling large amounts of language and keeping track of many contextual factors at the same time. It therefore has opened the way to a multitude of new investigations of language use” (as opposed to concentrating on language structure alone) (Biber, Conrad, & Reppen, 1998, p. 3). Further, it frees linguists from the bonds of “native-speaker intuition.” So many times our intuitions about language are wrong anyway. With corpus data, analysts no longer are limited to guessing what most people would say. Analysts can rely, with confidence, on empirical data—data which might show patterns and probably also shows language quirks that would have been overlooked without corpus data.

Biber, Conrad, & Reppen (1998) remind us, however, that corpus-based studies should not necessarily compete with other linguistic analysis approaches. Rather corpus approaches can work together with and complement those “more traditional approaches” (p. 9). Similarly, Stig Johansson (in Aijmer & Altenberg, 1991) notes that a corpus is only “*one* of the linguist’s tools, to be used together with introspection and elicitation techniques” and refers to Jan Svartvik as illustrating such methodology (p. 313). Aijmer and Altenberg identify corpus linguistic’s major potentials as 1) giving more realistic language descriptions, 2) providing great insights into language variation, or “language varieties,” and 3) “exploring the

quantitative and probabilistic aspects of language” (p. 2).

It is our hope that as the use of computerized corpora increases, the reliance on introspection and intuition so characteristic of much linguistic theorizing in the second half of the twentieth century will be balanced by more empirically grounded theorizing based on the facts of usage in English and every other language (Biber & Finegan, 1991, p. 220).

### **Register Variation**

One of corpus linguistic’s distinct advantages is its ability to give us insights into register variation. Biber & Finegan (1991) recognize that corpus linguistics enables us to investigate formerly unanswerable questions concerning text, genre, and register variation (p. 206). They point to the extensive potential of the Brown and LOB corpus and the London-Lund corpus, which contains numerous spoken-text registers with face-to-face conversation, telephone conversation, public interviews, panel discussions, radio broadcasts, parliamentary debates and speeches, court cases, dinner speeches, sermons and academic lectures (p. 209). Many researchers note, however, that this ability of corpus linguistics has not been fully exploited (Biber, 2001; Biber & Finegan, 1991; Conrad, 2000).

Biber, Conrad, and Reppen (1998) note that corpus linguistics makes systematic register analysis much easier than ever before (p. 137). They demonstrate four different types of corpus-based register investigations which occur at various levels of specificity: a particular grammatical feature (dependent clauses in spoken and written registers), general linguistic feature patterns in spoken and written registers, texts in academic disciplines, and internal sections of a single academic text.

### **This Present Study and Corpus Linguistics**

Although this study will not describe multiple characteristics of multiple registers, it will examine one particular language feature across two registers. This extensive comparison of reported speech across two registers, Conversation and News, will provide useful insights towards a more accurate description of reported speech in general.

## **The Big Picture: SLA Findings Related to Corpus Linguistics and Grammar Teaching: Implications for Reported Speech**

The field of second language acquisition (SLA) studies how people learn a second language. One of the specific areas within the field of SLA concerns itself with grammar. This area tries to answer questions such as: How do people learn the grammar of a second language? What teaching methods have been successful in teaching grammar? What sorts of grammar activities help students become proficient in a second language? What grammar teaching methods and activities do not seem to help students learn their second language?

Questions and issues such as these are inherently tied to my present research study. We need a more accurate description of naturally-occurring reported speech because SLA findings are currently showing that students need exposure to authentic, naturally-occurring language. Not only do they need *exposure* to the language used for real communication purposes that they will encounter in their daily lives, but they also need opportunities to *interact* with such language—to analyze it, form generalizations about it, recognize patterns within it, manipulate it, and practice it *in communicative contexts*. SLA research demonstrates that results from empirical corpus studies of grammatical topics can be put to good use; empirical results can be implemented in pedagogically sound, SLA-principled ways to facilitate student learning processes.

In the following sections I will briefly outline some of those SLA findings that concern possible implications for corpus linguistics and communicative grammar. In other words, I describe how the results of my present study could matter in terms of applying SLA findings to the teaching of grammar and presentation of reported speech. Specifically, it seems that corpus linguistics serves as an ideal tool for implementing SLA findings, particularly in three communicative context areas: Input, Focus on Form, and Consciousness Raising (C-R).



### **Input and Communicative Instruction**

Krashen (1985) gives prime importance to *comprehensible input*. Comprehensible input, or “i+1” (input + 1), emphasizes that people learn a second language by comprehending meanings in a second language. So mere exposure to a second language does not work; the exposure must be at least partly comprehensible to the learner. The portion that the learner understands serves as the stepping stone, “i,” for the new knowledge to be learned, “+1.” Exposing a student to input just slightly above her current level will cause her to rely on the portion of the language she comprehends and will enable her to learn the portion she does not understand, the i+1 theory claims. Krashen characterizes his Input Hypothesis as a not uncommon idea which he refined and named, but did not invent. Most simply stated by Krashen himself,

the input hypothesis claims that we acquire language in an amazingly simple way—when we understand messages. We have tried everything else—learning grammar rules, memorizing vocabulary, using expensive machinery, forms of group therapy, etc. What has escaped us all these years, however, is the one essential ingredient: comprehensible input (p. vii).

Nevertheless, Ellis (1994) claims that “no study to date has shown that comprehensible input enables learners to acquire grammatical features” (p. 604). And indeed, Krashen's claims remain largely theoretical.

Krashen's notions regarding input led, in part, to a movement in SLA theory toward communicative learning. Communicative learning can be seen either as a backlash to traditional, grammar-translation, memorization/drill-based language instruction, or merely as an attempt to recreate natural first language acquisition conditions. Communicative instruction aims to avoid explicit language (grammar) instruction, is more experiential than analytical, emphasizes negotiation of meaning<sup>1</sup>, and is largely an inductive process. These ideas relate to Krashen's SLA theory, and therefore view naturalistic, unstructured, meaningful language input as essential. In accordance with this, Krashen and other advocates of communicative approaches view grammar as explicit, conscious knowledge having no place in the classroom. That sort of explicit grammar is usurped by natural, spontaneous conversation.

---

<sup>1</sup> *Negotiation of meaning* occurs when people “struggle” to communicate.

In general, research into communicative classrooms has shown that a communicative focus provides definite advantages, but cannot be relied on, alone, to fully equip learners with all aspects of language learning. In other words, communicative instruction is good, but it cannot do it all. Ellis (1994) specifies that communicative classroom settings do not produce "worse" results than traditional classrooms, and that they "may help develop fluency and discourse skills," but do not ensure "high levels of linguistic and sociolinguistic competence" (p 604). In particular, communicative-only approaches tend to produce learners who lack grammatical accuracy! Along these same lines of criticism, Lightbown and Spada (1990) acknowledge that the French immersion program in Canada has been referred to (they cite Hammerly, 1987) as evidence for the ineffectiveness of communicative language teaching (p. 431).

Pica and Doughty (1985) also point to a need for grammatical input in addition to casual naturalistic input in order to combat "a stabilized nontarget variety of learner English" (p. 132). Most other "non-Krashen" SLA input researchers have tended not to refute the core idea of comprehensible input, yet have gone beyond " $i + 1$ " in their investigations of input to provide interesting insights. Ironically, Swain (1985) asserts that comprehensible input is "crucial" to acquiring grammar specifically *because* "it permits the learner to focus on form," even though Krashen maintains that the *form* of grammar should not be emphasized (p. 248). Swain's comments occur in the context of an argument for comprehensible output, yet she addresses comprehensible input again by noting that the value of comprehensible input will vary according to the type of input and specific area of SLA being acquired. Long's (1985) findings with foreign students in Hawaii also illustrate the notion of comprehensible input variance. His research "provides evidence of a causal relationship between linguistic and conversational adjustments of the kinds native speakers make to nonnatives under certain conditions and the comprehensibility of what they say to their nonnative listeners" (p. 388).

Collectively, then, communicative approaches seem to be beneficial in two formats: 1) alone, as the opposite alternative to non-engaging traditional grammar classrooms, and even better as 2) an implicit partner in combination with explicit grammar instruction. Numerous studies have indicated that a "form (explicit grammar) + meaning (communicative methods)" teaching approach works more optimally than either form alone or meaning alone. (See Fotos

& Ellis, 1991; Lightbown & Spada, 1990; Little, 1994; Spada, 1987) . Little (1994) argues that grammar instruction is appropriate within a communicative framework for three reasons: 1) explicit knowledge can contribute to and enhance implicit (grammar) knowledge, 2) explicit grammatical information is quite useful to learners when they are consciously needing and using such information, as in a situation where Krashen's monitor theory would be applicable, and 3) explicit grammatical knowledge can be an advantage in certain linguistic situations when a linguistic gap occurs (pp. 103-105).

Ellis (1994) synthesizes the SLA input research, though filled with mixed results and sometimes methodologically flawed, as generally indicating a correlation between input frequency and output accuracy of specific grammatical morphemes (p. 286). Despite the variance in their research conclusions, Ellis (1994) notes that the empirical studies about input have in common the finding that input *is related* to acquisition (p. 268). We can conclude that input does affect acquisition, even though a significant portion of the specific means and details of *how* input affects acquisition are unknown. Of course the degree to which it influences acquisition varies, but the fact remains that input, including Krashen's comprehensible input, is a factor in SLA.

### **How input relates to this study**

Given the nature of SLA input, corpus-based techniques serve as excellent tools for learners to gain access to huge amounts of quality, naturally-occurring, native-speaker data. Additionally, corpus techniques provide an ideal setting for enhancing and modifying data—i.e., transforming input into comprehensible input. This study's empirical corpus findings about reported speech can provide both the input *and* the grammatical guidance that SLA research shows as optimal. Reported speech corpus data provides abundant material to be tailored into comprehensible input. Instructors can manipulate this corpus data to address particular grammatical forms in communicative contexts. This is extremely useful for SLA-guided teaching purposes such as consciousness-raising, noticing, and focus on form, which will be discussed next.

## **Focus on Form**

### **Focus on form defined**

Focus on form involves drawing learners' attention to linguistic forms in the context of a communicative or content-based setting. Focus on form merges explicit grammar instruction with communicative language instruction. Therefore, focus on form coincides with the research about input (above) that found input most successful when combined with explicit grammar.

This *focus on form* should be contrasted with *focus on formS*. Krashen's ideas were partly a reaction against the prescriptive-grammar-based, repetitive drilling and memorization language teaching tradition: a focus on formS approach. Long and Robinson (1998) distinguish the two approaches to SLA grammar acquisition as "analytic" and "synthetic" respectively (p. 16). Focus on formS, however, is driven by the "teaching and testing of isolated linguistic units" (Long, 1991, p. 41). Teachers often implement Focus on formS by using a sequential grammar syllabus to guide their class (week 1 = plural 's'; week 2 = third person singular; week 3 = irregular verbs). Focus on form, however, is implemented through *meaning*. Focus on form uses language communicatively to express meaning, and will draw attention to grammatical form as needed—in the context of relevant, meaningful communication. The focus on form approach thereby merges a communicative approach to SLA input (which, in pure form, would exclude any grammatical form elements) with the more traditional notions of formal grammar being central to second language acquisition.

So focus on form differs from traditionally rigid, decontextualized grammatical drills. Focus on form applies grammar points in a context—a natural context—in a communicative setting. Grammar, in focus on form, occurs in context of meaningful interactions; focus on form actually balances the extremes between grammar-only (focus on formS) and natural-conversation-only (Krashen's naturalistic acquisition) approaches. Therefore, focus on form becomes a more reasonable compromise between structured and free, grammar and natural.

### **Focus on form evidence**

Research into focus on form has come from diverse areas of SLA and has been

motivated by various causes. Some research has been highly theoretical, while some has been more practically oriented. Some key pieces of research which highlight the major evidence supporting focus on form include Doughty & Varela (1998), Fotos & Ellis (1991), Lightbown & Spada (1993, 1990, 1987), Pica (1994), Spada (1987, 1990), and Swain (1998). Pica (1994) concludes that, in many situations, learners need more than just input and interactional experiences alone; they also need some form of explicit linguistic instruction for the sake of efficiency. Pica calls for "a balance in the classroom between explicit instruction and more inductive, communicative procedure," and cites another focus on form study (Montgomery and Eisenstein, 1986) with the same results as Lightbown, Spada, and most other focus on form researchers: that students who had access to both grammatical form instruction *and* communicative experiences "showed the strongest gains" (p. 67).

Criticism of focus on form has been limited to VanPatten (1988) and Cook (1999). VanPatten asserts that evidence shows focus on form to not be beneficial. Yet VanPatten seems to have misinterpreted the notion of grammar instruction in focus on form as more akin to focus on formS. Cook (1999) claims that linguistic forms are too complex and misrepresented to describe to students.

Corpus-based focus on form techniques, however, counteract Cook's and others' criticisms. Corpus data can actually *show* learners how particular grammatical features are used and let learners discover the patterns themselves. Partington (1998) also tempers Cook's claims by reminding that [form-focused] corpus techniques should be used "not as a surrogate *teacher*, but as an informant or a pedagogue or *educational slave*, in other words as a resource of information to be tapped by the learner" (p. 5). Therefore, any cautions against form-heavy information overload should realize that a good teacher would implement corpus techniques in effective, non-overwhelming, relevant ways. Similarly, Conrad (2000) suggests samples of these methods by noting corpus techniques' successes with focus on form in specific sociolinguistic situations, pragmatic contexts, and register variations.

### **How focus on form relates to this study**

As with input, corpus linguistics serves as an ideal tool with which to implement focus on form. Again, corpus studies provide abundant form-focusing data for students to work

with, and corpus data can be easily manipulated to emphasize particular grammatical features. Reported speech grammatical data from this study can provide data for the form focused instruction that SLA research advocates.

## **Consciousness Raising**

### **Consciousness raising defined**

One aspect of focus on form, consciousness raising (C-R), has become quite prominent. Consciousness raising, in line with focus on form, involves drawing learners' attention to particular grammatical features, but not necessitating explicit instruction, nor requiring any type of learner production of these linguistic principles. So a teacher might teach a communicative lesson, but at some time within that lesson make students aware, at a conscious level, of a specific grammatical point. This could involve highlighting the grammatical feature in the instruction materials, pointing out the feature each time it occurs, asking students to notice and look for the feature—maybe count each time it occurs, record what language surrounds the feature, note what part of the sentence the feature tends to appear in, etc.

Rutherford (1987a) emphasizes that C-R should *not* consist of an accumulation of separate language rules (decontextualized, discrete forms), but that grammar should incorporate a product-based approach. Rutherford argues that C-R should reflect the principles of the "nature of language" and the "nature of language acquisition" (p. 209). Rutherford advises researchers to "consider C-R with regard to grammatical processes and the learner's progressive restructuring of prior knowledge" (p. 209).

Research in the area of C-R indicates that a C-R approach to form-focused grammatical SLA has been and tends to be successful. Rutherford (1987b) connects C-R theory to related research issues as justification for C-R in SLA. Rutherford reiterates familiar themes, noting Harley and Swain's (1984) findings that comprehensible input was not enough, Spada's (1987) assertions that learners need both form-focused and function-focused practice, and Pica's (1994) findings of how classroom instruction positively affects learner hypothesis forming. Yip (1994) also supports C-R on grounds that it is "a middle ground" and that one of

her studies shows empirical success (p. 124). Yip demonstrates how C-R can effectively alleviate some of the difficulties associated with ergative verbs, though her study was plagued by the routine "learner-centered variables" as well as particular factors of motivation in some specific learner cases (p. 136). Ellis (1994) also supports methods of consciousness raising for communicative grammar instruction.

### **How consciousness raising relates to this study**

The very format of corpus data lends itself to consciousness raising. The basic layout of this study's corpus data automatically highlights reported speech, repeatedly, as a specific language feature (see the KWIC samples in Appendix A). Corpus data naturally draws students' attention to individual language features while also providing plenty of context around the feature. This study's corpus data can be discussed communicatively; the highlighted language feature does not need to be taught explicitly, but the reported speech feature will be noticed by students exposed to the corpus results.

### **SLA Findings Related to This Present Study**

This brings us to the "so what?" of these SLA findings. Since SLA research indicates that a communicative focus on form approach to grammar does work best, what does this mean for the application of this study's findings?

*The very nature of corpus linguistics lends itself to an application of SLA grammar research.* In fact, corpus linguistics seems to be uniquely ideal, in numerous areas, to facilitate SLA grammar findings—including grammar findings about reported speech. Teachers and materials writers can use corpus study results to implement focus on form, consciousness raising, and other communicative input techniques. By and large, most research into corpus technique applications have reported and predicted promising successes (Conrad 1999, 2000; Fox, 1998; Johns, 1994; Partington, 1998; Stubbs, 1996). Fox claims that corpus-based activities will help instill in students an awareness and curiosity about language, which will also help develop analytical skills (p. 42). These assertions align with communicative notions of focus on form and suggest classroom activities and materials including frequency information, contexts and co-texts, grammar noticing and focusing, collocation and

phraseology, pragmatics, and other concordancing uses.

I have cited several representative illustrations of research trends into SLA grammar. Their natural affinity relationship with corpus linguistics, given recent technological abilities, should be apparent. Since SLA research points to trends towards communicative form-focused grammar instruction, balancing implicit and explicit instruction, and incorporating elements of consciousness raising, noticing, and comprehensible input, corpus linguistics should be recognized as an ideal tool to implement these SLA findings. Not only do corpora allow for principled SLA grammar instruction, but they also give a high element of flexibility to practitioners. Teachers, as well as researchers, can select relevant grammatical information at will from a large corpus of data and be more assured of its validity and authenticity. Ignoring this powerful potential would sincerely inhibit the relevance of SLA grammar findings.

## Summary

This chapter has covered research literature about reported speech, corpus linguistics, and second language acquisition. It has argued that an empirical corpus study of reported speech is needed to discover valuable information about naturally-occurring reported speech behavior. It has also argued that results from corpus studies can greatly benefit students' SLA processes. Specifically, results from this corpus study can help ESL/EFL students master the grammatical topic of reported speech. SLA research has found a form-focused, meaning-based approach to grammar, in conjunction with a communicative approach, to be most beneficial to students.

With this theoretical background covered, we now begin a more concrete discussion of the present corpus study itself. Chapter three will present the methodology of my empirical reported speech research.



## CHAPTER 3. METHODOLOGY

This chapter presents the methods of gathering and analyzing empirical data for both the textbook study and the corpus study. This chapter also explains decisions made throughout the analysis process as well as describing particular problems encountered.

I first explain how I determined what to consider as reported speech. I then explain the textbook study in terms of its data collection, data description, and data analysis processes. I explain the data collection for the corpus study and explain how I analyzed the corpus data by finding the reported speech samples, calculating proportions and frequency counts, and labeling verb tense combinations.

### Identifying Reported Speech

The first step was to devise a reliable system of identifying reported speech. This was important for both the corpus and textbook analyses for uniformity purposes, since both analyses needed access to the same definition. Through an analysis of grammar textbooks, research articles, and sample positive and negative examples, I arrived at the following informal definition: Reported Speech is merely the language utterance that occurs when someone needs to explain what someone (or something) has said, but without repeating it exactly—i.e., *not* direct speech or something that should appear in quotation marks. Reported speech must contain two verbs as well: a main (reporting) verb and an embedded “clausal verb,” as in “He said he was sick” or “She says she’s hungry” or “They told us they would stop by tonight.” Using these defining characteristics, it is possible to distinguish between reported speech and other constructions (constructions that are similar, but are not technically reported speech):

(The register that the corpus samples were taken from follow each item)

## Reported Speech:

- *The engineer said his family would harvest four sacks and give two back to the collective farm* (news)
- *They say he works out of his home* (news)
- *and my uncle said he'll, you know, send me money every once in awhile* (conv)
- *I told him I was traveling with my fiance who is gonna be my wife* (conv)
- *and I told him inside that I will do everything I can to generate support* (news)
- *he told Republican audiences he would take his case against Democrats* (news)
- *Another recent study tells us that 70 percent of Americans get 100 percent of their information from TV* (news)

## NOT Reported Speech:

- *one bank told her, "no woman should get more than \$50,000"* (news)  
[this is direct speech]
- *make sure you tell Ed that I want my pictures* (conv)  
[this is a command/request to say something in the future, not a report of past speech]
- *tell me, uh, do you know Professor Jones?* (conv)  
[a figure of speech using "tell" to introduce a regular question—not reported speech]
- *I'm wondering if he really told his father the truth* (conv)  
[a present act—"I am wondering"—not a report of a past act or statement]

Some samples are ambiguous, but certain factors can help distinguish tricky examples. Pronoun shifts can be a key.<sup>2</sup> Imperative commands for the future can also rule out some confusion since such a construction cannot be reporting a prior statement or action. A reporting verb followed by a direct object with no embedded clause such as "tell the truth" or "he tells us what we don't know about Eisner" is not reported speech either.

Yet so much of language depends on context, so numerous ambiguities surface when

<sup>2</sup> For instance, though "John said I'm sick; he has a terrible headache" could potentially be reported speech given the lack of quotation marks (with John originally saying, "you're sick" and meaning that the other person—not him—feels ill), we can tell that this is not reported speech since the pronoun shift would not make sense given the contextual information (John is the one who is sick because John is the one who has a headache). To be reported speech in this context, the statement would have to read "John said he's sick; he has a terrible headache" If a reporter says the first version, "John said I'm sick. He has a terrible headache," the reporter's voice probably changes pitch on the "I'm sick" portion in order to indicate that the reporter has switched to direct speech and is now mimicking John's actual words by pretending, momentarily, to be John and act in the role of John as if on stage. This would be similar to the colloquial use of "like": "... and John was like, 'I'm sick' ... and I was like, 'oh that's too bad' ..."

looking through somewhat decontextualized corpus samples. For example:

- *And he said I'll harvest four sacks and give two of them back.* (this is ambiguous since no quotation marks; it could be interpreted as reported speech or direct speech: "he said, 'I'll harvest four sacks'..." OR "he said that I'll harvest four sacks...")
- *then a neighbor told us it was our water softener* (news) (again, this could be either "a neighbor told us that it was our water softener" OR "a neighbor told us, 'it was our water softener'")

Despite the variety of ambiguous samples, having one uniform definition of reported speech to apply makes the decision process easier. Having established the characteristics of reported speech identification in the beginning, the decisions about reported speech samples are more consistent and reliable. The majority of samples were not ambiguous, and those that were ambiguous often contained other clues that made their identification more reliable.

## **The Textbook Analysis**

The textbook analysis gave me information about what ESL grammar textbooks tell students about reported speech. It also helped guide the corpus study by setting the initial search parameters to search through the corpus with.

### **Data Collection**

ESL/EFL grammar textbooks were chosen from the Intensive English program's library at Iowa State University. The books were kept as references for this large Midwestern university's intensive English program. Many of the books in the library had been used as classroom texts; instructors also used many of the books as supplementary material in their classrooms. I looked through all books in the "coursebook," "composition/writing," and "grammar" sections of the library and found 19 books that discussed reported speech. After examining the 19 books, a decision to focus on EFL/ESL grammar textbooks was based, primarily, on the fact that both the coursebooks and composition books provided fairly sparse coverage of reported speech. The grammar book explanations were much more thorough and

provided better quality information at the level of detail I needed for the corpus study.

From the grammar books, seven were chosen for this study based on the amount of reported speech coverage, the date published, and the publisher. I looked for grammar books that had been published fairly recently and that actually did cover reported speech in a meaningful way. That is, the books needed to devote a substantial section to the topic of reported speech (a few pages, a small unit, a major portion of a chapter)—something more than a casual mention or a few brief examples with no discussion. Seven grammar textbooks were found to fit the criteria. This means that these seven fairly well-known ESL grammar books selected contained the most thorough, complete coverage of reported speech available. The other available books examined did not have such level of detail.

### **Description of Data**

The seven grammar books chosen came from a variety of publishers, were recently published, and seemed to be fairly well-known in that they tended to be typical grammar references for ESL instruction; they were kept as resources in Iowa State's intensive English program, sometimes mentioned in ESL-teaching discussion-group e-mails that I had read, and often referred to in informal discussions among ESL teachers at conferences I had attended. As mentioned, they all devoted at least one section of their grammar topics to reported speech; this means I found a unit or chapter of substantial reported speech coverage—something I did not find in most of the other books I looked through. The amount of pages they used to discuss reported speech ranged from four to more than twenty. All seven of the books were published in the last 15 years (one in 1989, the rest in 1995 or later), and they represented a variety of major ESL/EFL publishers.

The study used these seven ESL grammar textbooks:

- Bland, S. K (1996). Intermediate grammar: From form to meaning and use (pp. 461-475). Oxford: Oxford.
- Eastwood, J. (1999). Oxford practice grammar with answers (pp. 318-329). Hong Kong: Oxford University Press.
- Elbaum, S. N. (2001). Grammar in context (3rd ed.) (pp. 362-385). Boston: Heinle & Heinle.
- Fuchs, M. & Bonner, M. (1995). Focus on grammar: A high-intermediate course for

reference and practice (pp. 271-328). Addison-Wesley; Longman.

- Murphy, R., Altman, R. & Rutherford, W. E. (1989). Grammar in use: Reference and practice for intermediate students of English (pp. 90-93). New York: Cambridge University Press.
- Raimes, A. (1998). Grammar troublespots: An editing guide for students (2nd ed.) (pp. 141-145). New York: Cambridge University Press.
- Thewlis, S. H. (2001). Grammar dimensions (3rd ed.) (pp. 400-416). Boston: Heinle & Heinle.

These grammar textbooks all stated explicitly that they were written for ESL/EFL students. Some grammar books referred to reported speech as “indirect speech.” Among the grammar books was a wide variety in the amount of coverage and method of presenting reported speech, but most all of them included practice exercises of some sort.

## **Data Analysis**

### **Focus**

I limited the *focus* of the textbook analysis to three aspects: 1) which verbs the textbooks presented as reporting verbs, 2) which tense combinations the textbooks presented as appropriate, and 3) what other rules the textbooks gave about reported speech. The first two focus points addressed the usual controversial points that textbooks were criticized for. The last focus point was intended to give a more complete picture of textbooks’ total reported speech presentations. I also noted all aspects of each grammar book’s reported speech presentation including the topic’s prominence in the textbook, informational statements made, samples provided, cartoons used, practice exercises, etc.

These three focus points become evident in this thesis’ chapter four results, as the textbook analysis discussion centers around these three components. This more limited focus also helped guide the corpus study.

### **Analyses**

In accordance with the focus points, I wrote down detailed notes about each textbooks’ reported speech coverage. I noted the reporting verbs it presented and how it explained the

tense combinations (and backshifting rule). I also recorded any other directions (explicit rules) the textbook gave concerning reported speech. This compiled information enabled a standard comparison across textbooks to take place. The information also showed common patterns among the textbooks, and from this profile, I made a general assessment of ESL/EFL grammar textbook coverage of reported speech, the details of which are discussed in chapter four.

## **The Corpus Analysis**

The empirical corpus study was set up to provide a description of how reported speech is actually used in authentic language. In order to search for and isolate occurrences of this naturally-occurring reported speech within the corpus, the search parameters had to initially rely upon traditional assumptions—those outlined in grammar books—and other intuitions about reported speech behavior. For instance, the reporting verbs used in the computer's search algorithm were the reporting verbs put forth by the grammar textbooks. Since the grammar textbooks had a prominent role in guiding the corpus study, the empirical corpus results themselves serve as a sort of “check” on the accuracy, or at least the authenticity, of grammar textbooks.

### **Data Collection**

#### **Selection of the corpus**

The corpus analysis used two registers from the Longman Spoken & Written English corpus: American Conversation and American Newspapers. The American News register totals ~3.3 million words, and the American Conversation register includes ~2.1 million words total. To put these word counts in perspective, one million words is estimated to equal ~4,000 double-spaced manuscript pages, or 2,850 book pages, or 140-150 hours of conversation (spoken at an average rate of ~400 words/minute) (Biber, Johansson, Leech, Conrad, & Finegan, 1999, p. 39). The News register contains data from all sections in newspapers

including business, entertainment, politics, world news, and editorial topics. The corpus' newspaper content comes from the *AP wire service*, *The Wall Street Journal*, and *The San Jose Mercury*. The corpus' conversation data comes from 491 speakers (199 male; 292 female) who ranged in age from teenage to over 60 years old. The speakers various U.S. geographic regions, educational levels, and ethnic backgrounds (Biber, Johansson, Leech, Conrad, & Finegan, 1999, p. 32).

These two specific registers, Newspapers and Conversation, were chosen for this study to complement one another and set up an interesting comparison between two distinctive types of language. The Conversation register represents an informal, casual, everyday type of spoken discourse while the Newspaper register contains a more formal type of written discourse. Additionally, the Newspaper register was used for this study because ESL/EFL students find News to be a particularly relevant register. Students usually encounter News in their everyday lives, and so it is a very accessible register to students for receptive practice. And to some extent, interpreting information in news reports is a basic life skill that non-native-speaking students need to become proficient in. Also, because of the nature of news and its purpose to report information, one would also expect a News register to use lots of reported speech. Granted, it is possible that the news sources used—particularly the AP wire service—have particular style guidelines that might include rules about reporting speech; perhaps all of the AP's published articles are required to conform to a particular formatting style. Yet AP wire service articles are so widespread throughout all types of U.S. newspapers that any AP-mandated-style would be completely appropriate in a description of the American Newspaper register. Because AP articles are so prevalent, their stylistic influence seems inevitable.

This corpus used in the study is large enough and principled enough to be representative. Of course, the corpus does not represent all possible naturally-occurring language—a corpus is only as representative as its contents. Specifically, this corpus used represents the particular varieties of language it contains: spoken American informal conversation and written American newspapers (all sections). Yet the Longman Spoken & Written English corpus, from which this study's smaller corpus draws two registers from, was systematically built by corpus linguists and has been used in many reputable research projects.



The latest and most comprehensive English grammar description, The Longman Grammar of Spoken and Written English (Biber, Johansson, Leech, Conrad, & Finegan, 1999) used this Longman corpus.

### **Searching the corpus with computer programs**

Computer programs (written by Conrad) used a search algorithm to try to find all possible reported speech occurrences in the corpus. The search algorithm used the reporting verbs that textbooks had identified: *say*, *tell*, and 43 “other” reporting verbs—*add*, *admit*, *advise*, *announce*, *answer*, *assure*, *ask*, *claim*, *comment*, *complain*, *confess*, *convince*, *declare*, *demand*, *explain*, *indicate*, *inform*, *insist*, *invite*, *knew*, *mention*, *notify*, *order*, *persuade*, *point out*, *promise*, *propose*, *recommend*, *remark*, *remind*, *reply*, *replied that*, *report*, *require*, *shout*, *state*, *suggest*, *swear*, *teach*, *think*, *want to know*, *warn*, *whisper*. Occurrences of *say* and *tell* included “say,” “says,” “said,” “saying,” and “tell,” “tells,” “told,” “telling,” and the computer program also searched for the various forms of the other verbs. *Say* and *tell* were the main verbs that grammar textbooks identified with reported speech, but the textbooks had identified the “other-non-say-tell” verbs as additional reporting verbs besides *say* and *tell*. So the computer program was also searching to see how commonly those other textbook verbs occurred.

The search algorithm looked for reporting verbs that were in a grammatical position likely to be reported speech. The possible reporting verbs (*say*, *tell*, and the others) were coded into the program, and the algorithm isolated all occurrences of the verbs and then judged whether or not the surrounding context was probable reported speech grammar. For instance, if the algorithm located the sentence “*The president was upset, White House officials said.*”, it would know that the occurrence was not reported speech since the reporting verb, *said*, was at the end of the sentence. The output results were designed to follow the particular grammatical patterns that reported speech follows: ‘subject’ + ‘reporting verb’ + ‘embedded clause with a subject+verb’. To reach this point, the search algorithm was improved through previous pilot runs.

Therefore, the nature of this data collection aimed to produce output results containing reporting verbs that were in a likely environment and position to be reported speech. The



output results were also designed to avoid occurrences that were not reported speech. However, given the complexity of this linguistic data, an automatic computer analysis alone could not possibly filter out all non-reported-speech samples from the output with 100% accuracy. Some of the output occurrences were other language instances that happened to fit the reported speech pattern grammatically, but were not actually reported speech. Therefore, additional hand-editing was necessary.

### Hand editing of data

The computer program's search algorithm produced large amounts of output data—data that contained a reporting verb in a position likely to be reported speech. This output data was in KWIC (Key Word in Context) format, and so could be viewed with a simple word processing program (see the sample below). KWIC format is simply a way that the computer displays data from the corpus in a format that is easy for a person to read. In KWIC format, the word being searched for is surrounded by extra blank space and then whatever amount of language context the programmer specifies. (More sample KWIC files can be viewed in Appendix A.)

#### Sample KWIC format:

```

                        Item = told
                        File = 001-0116.AP
in the well - to - do birmingham suburb because
ms . watson had taken it with her . neighbors
-----> told
the Post - herald that Deneau was a stockbroker
who had been unemployed for more than a year and
that he was intensely private , posting `` no
trespassing " signs to keep people away from his yard

                        Item = told
                        File = 001-0174.AP
components in a $ 4.7 billion deal . democratic sens
. alan dixon of illinois and wendell ford of kentucky
-----> told
reporters they plan to introduce legislation requiring President Bush
to give House and Senate leaders documents related to the
U.s. - south Korean deal worked out last year .
`` to intelligently exercise our constitutional authority , we
```

As explained previously, I could not assume that all the output occurrences were reported speech. But the occurrences could be thought of as potential reported speech instances because they had been screened through the computer program's search algorithm. Therefore, hand-editing was necessary to sort out the real reported speech occurrences from the potential reported speech occurrences.

The tedious process of hand-editing involved going through the filtered files to extract true reported speech occurrences. Since the four main *say-tell* files were so extensive, I looked at representative samples. For example, from the News register, the computer program returned as output 17,961 occurrences of the verb *say* that were potentially reported speech. Displaying this output data required almost 3,600 pages! I could not feasibly check all 17,961 occurrences. Instead, I followed a uniform sampling procedure to go through the output results of each verb in each register. Beginning with the first output occurrence, I looked at every 100th occurrence until the end of the corpus. So for the *say* occurrences in News, this produced ~180 sample occurrences. I then repeated the process twice (the process of looking at every 100th output occurrence), but starting with the 50th and 25th occurrence instead of the first occurrence in order to spread the sampling throughout the corpus. The two repetitions produced two more data samples of ~180 occurrences each. So, using *say* in the News register as an example, by the end of the sampling procedure, I had looked at 540 occurrences of *say* in News.

These 540 output data occurrences were used as one of the representative samples in the quantitative corpus data analyses. I classified each of the 540 occurrences as either reported speech or not reported speech according to the previously-explained reported speech identification process. To check the representativeness of the entire 540-occurrence sample, I compared the classification results of each of its 180-occurrence subsamples to ensure that each subsample produced uniform proportional results. In other words, the first subsample (every 100th occurrence beginning with #1) produced 138 positive and 41 negative reported speech occurrences; the second subsample (every 100th occurrence beginning with #50) produced 134 positive and 46 negative; the third subsample produced 134 positive and 46 negative. Since all three proportions of its subsamples were quite similar, I could assume that the results from the combined 540-occurrence-sample were trustworthy. (See Biber, Conrad, & Reppen, 1998,

pp. 87-93 for another example and explanation of this sampling procedure.) The same sampling procedure was followed for the output data of *tell* in both registers. No other output data file was as large as *say* in News, but the same proportion of data was looked at for each representative sample (*say* and *tell* in both News and Conv). Table 3.1 below shows the numerical results of these representative samplings, and Appendix C shows more detailed results from the output data.

Table 3.1. Data from the Sampling Procedure

	Say-NEWS	Say-CONV	Tell-NEWS	Tell-CONV
# of occurrences returned as output from computer's search algorithm	17961	4222	1570	1479
# of representative occurrences looked at for sampling	540	133	48	45
# of reported speech occurrences in the sample	413	66	34	20
# of NON-reported speech occurrences in the sample	127	67	14	15
Proportional rate of reported speech	76.50%	49.60%	70.80%	44.40%
Frequency rate of reported speech (proportion rate normed according to corpus word count)	4205/million	1011/mill.	340/mill.	317/mill.

Since the sampling procedure with the three other major files (*say-Conv*, *tell-News*, *tell-Conv*) produced much less data than *say-News* (413 positive reported speech samples), after the representative samples had been gathered and numerically recorded (positive reported speech: *say-Conv*: 66; *tell-News*: 34; *tell-Conv*: 20), more positive reported speech samples were gathered randomly in order to have ~100 samples to work with for the verb tense analysis. Similarly, because the other non-*say-tell* verbs produced even smaller amounts of data, larger proportions of their output were looked at—I looked at every example in most cases; for a few of the non-*say-tell* verbs that had large amounts of data, I looked through representative samples of every few pages. See Appendix C for exact details of the “other verb” data analysis. As with the larger output files, keeping track of how many occurrences I classified as reported speech throughout the process enabled me to tally frequency rates.

## **Data Analysis**

Two types of quantitative results were figured for the reported speech data analysis: *proportional rates* and *frequency rates*. Most of the reported speech analysis (discussed in chapter four) uses the frequency data. However, reported speech proportional data is needed to determine reported speech frequency data. Additionally, the proportional data itself becomes useful when analyzing the smaller non-*say-tell* verbs (see chapter four's discussion and Appendix C). Both proportional and frequency data are explained below.

### **Reporting verb frequency calculations**

Within the filtered files, all occurrences were potentially reported speech. Of course, no file had a 100% accuracy rate of occurrences actually being reported speech, though. So the *proportion* of actual reported speech in a file needed to be determined by simply dividing the number of actual reported speech instances from the representative sample by the total number of potential instances looked at in the file's representative sample. So the *say* in the News register had a 76.5% proportional rate since 413 of the 540 representative samples had actually been reported speech (127 had been judged as not actually reported speech). As previously mentioned, the total 540-occurrence sample's 76.5% proportional rate correlated with all three of its subsample proportional rates; the three 180-occurrence subsamples all had proportional rates of ~75%. Therefore, the 540-occurrence sample's proportion could be trusted to accurately represent reported speech behavior using *say* in News. With the three smaller files, (*say* in Conversation, *tell* in Conversation, and *tell* in News) I used the initial numbers from the first representative samplings and did not include the less systematic counts I had had to obtain when gathering the extra samples to reach the 100 occurrence level. (So *say* in Conversation used 66 [not 100] positive out of 133 total; *tell* in News used 34 [not 100] positive of 48 total; *tell* in Conversation used 20 [not 100] positive of 45 total).

These proportions were then normed to one million words to determine the reported speech *frequency rates*. This was intended to give a more accurate measure of comparison between the registers. The process of determining verb frequencies by norming to one million occurs by dividing the total number of positive reported speech occurrences in the data sample by the total word count of the Register and then multiplying by one million. In this way, the

frequency count for each verb in each register was determined.

### **Analysis of verb tense combinations**

The final portion of the data analysis involved going through each *say-tell* reported speech sample and classifying the verb tense combination. This was necessary to determine whether or not the reported speech had been backshifted according to textbook prescriptions. I considered the verb tense of both the main (reporting) verb and the embedded clausal verb. Each verb was considered on two levels: a detailed tense-labeling level and a broad, general level. The general level consisted of two options: Past or Present; the detailed level consisted of all other possible verb tense permutations such as past perfect, present progressive, passive, present modal, etc. (which accounted for verb tense, aspect, voice). Therefore, the reporting verb of “*The president has said he will reduce taxes*” would be labeled at the detailed level as present perfect tense and labeled at the general level as present tense. The embedded clausal verb would be labeled at the detailed level as present-Modal+present (M-present + present) and labeled at the general level as present. The specific verb tense combination categories are shown in Appendix B.

Once both verbs in a reported speech occurrence had been labeled, the tense combination became evident. The detailed tense labels determined the more general tense classifications. On a basic, broad level, verb tense combinations could be Past-Past, Past-Present, Present-Past, or Present-Present. So, for instance, “*He said he had been sick*” would be labeled past--> past perfect on a detailed level. The past perfect tense is categorized under the general tense of “past,” and so at a general level, the sentence is Past-->Past. Likewise, “*He wasn't saying that I had been wrong*” would be labeled as a past progressive --> past perfect detailed tense combination, and then categorized as a Past-->Past general tense combination.

Modal verbs and forms of “do,” which occurred in conjunction with another verb, would often determine the general tense combination. I would note both the form of the modal or “do” verb along with the tense of the neighboring verb, and use these distinctions in the detailed verb combinations classification scheme (see Appendix B). Although modal verbs technically do not have tense, they do exhibit distinct “forms” which I referred to as past (as in

*could*) and present (as in *can*). For example “He said that he couldn’t love her” classifies as a past -->M-past+present detailed combination, which translates to a Past-Past general combination. “Bush said Strauss will make sure that two ships will not pass in the night ...” (*news*) classifies as a past-->M-present+present detailed combination, which translates to a Past-Present general combination. Concerning “do,” “he said he didn’t immediately realize the ship had grounded” (*news*) would classify as a past-->DID+present detailed combination, which translates to a Past-->Past general combination.

The frequencies of each tense combination were recorded—at both the general and detailed level. This resulted in proportional verb tense data for each register. The proportion of Past-Past, Past-Present, Present-Past, and Present-Present combinations in each register were calculated. Additionally, the proportion of detailed combinations within each general combination was calculated for each register.

### **Comparing the Textbook Findings and the Corpus Findings**

Textbook assertions about reported speech behavior were compared with empirical corpus results on two main levels:

- 1) Reporting Verb Frequencies — Textbook claims about *say*, *tell*, and other reporting verbs were checked with corpus data to see if they matched.
- 2) Verb Tense Combinations (backshifting) — The corpus data’s verb tense combination information was compared with textbook directives to backshift verb tenses. The corpus data served as a check to see if and how often naturally-occurring reported speech followed textbook backshifting instructions.

## **Summary**

This chapter presented 1) the ways that data were collected, 2) the methodology for analyzing the textbooks’ treatments of reported speech, and 3) methodology for analyzing the corpus. Chapter four will present the results of these methods.

## CHAPTER 4. RESULTS AND DISCUSSION

This chapter presents the heart of the study. The results of the study are organized according to A) Quantitative Findings and B) Interpretations/Discussion. The quantitative findings section has two subsections. The first, Textbook Findings, categorizes results according to [a] reporting verbs used in textbooks, [b] tense combinations presented in textbooks, and [c] rules explicitly given in textbooks. The second subsection, Empirical Corpus Data Findings, categorizes results according to [a] reported speech frequency rates, [b] potential student problems, [c] verb occurrences, [d] tense combinations, [e] register variation, and [f] influential conditions. The discussion section is organized in terms of 1) Research Question Answers and 2) General Implications, including textbook v. corpora specifics and SLA-principled grammar teaching.

### Quantitative Findings

#### Textbook Findings

The textbook study results describe a general consensus about reported speech behavior as depicted in seven representative ESL/EFL grammar textbooks. In other words, most ESL/EFL students relying on textbook information would likely be given this information about reported speech. The information divides into three categories: 1) *verbs used*—which verbs are presented as “reported speech verbs;” 2) *tenses used*—verb tense combinations between the reporting verb and the embedded clause verb; 3) *rules*—rules that the textbooks explicitly give students regarding reported speech.

#### **Reporting verbs in textbooks**

*Say* and *Tell*, as expected, are the most commonly presented reporting verbs in textbooks. Virtually all textbooks in the sample focused on *say* and *tell*. The textbooks also

mention a wide variety of additional reporting verbs, sometimes categorizing the other verbs according to whether they behave like *say* or like *tell*, but the textbooks use *say* and *tell* most often in examples and exercises. One textbook even stated outright that *say* and *tell* are the most common reporting verbs (Eastwood, 1999).

Besides giving priority to *say* and *tell*, most of the textbooks also distinguish behavior differences between the two verbs. Elbaum (2001), explains that we *say* something, but we *tell* someone something. Or, as Murphy, Altman, and Rutherford (1989) say, we can't say "Tom \*said us about his trip to Mexico" (p. 92). Rather, if we say who we are talking to, we use some form of *tell*; otherwise, we use a form of *say*. In this way, most textbooks emphasize the presence of an indirect object in reported speech clauses which use *tell*.

It is according to this same *say-tell* behavior distinction that additional reporting verbs are often presented. The additional verbs mentioned, however, vary widely among the textbooks. There seems to be no general consensus, nor any sort of composite "other-reporting-verbs" list. One text (Bland, 1996) classified the additional verbs as "communication verbs," but, in general, textbooks identify a rather random assortment of "other" verbs as reported speech candidates. Table 4.1 (below) illustrates the variety.

In summary, the overall textbook treatment of reported speech verbs exhibits some common themes as well as individual idiosyncrasies among specific textbooks.

*The General Consensus:* *Say* and *Tell* are the most common reporting verbs; they behave differently in terms of whether an indirect object is required.

*Other Specifics also mentioned by one or a few textbooks:* Various textbooks list a variety of other reporting verbs, and sometimes they are not consistent in terms of *say-tell* classification; one textbook does not list any other verbs specifically (Murphy, 1989).

## **Verb tenses in textbooks**

*The embedded clausal verbs within the reported speech* All of the grammar textbooks surveyed instructed students to shift verbs within the reported speech clause to the past tense, except in some specific cases. So, *Julie*: "Thomas is sick" would become



Table 4.1. Reporting Verbs Given in Textbooks

Grammar Textbook	Other SAY-type Verbs	Other TELL-type Verbs	Others - No Distinction
<i>Bland (1996) Intermediate Grammar</i>	admit, announce, comment, complain, confess, explain, indicate, mention, point out, remark, reply, report, shout, state, swear, whisper	assure, convince, inform, notify, persuade, remind	advise, answer, asked, demand, insist, promise, propose, recommend, require, suggest, want to know
<i>Eastwood (1999) Oxford Practice Grammar</i>			announce, answer, explain, knew, mention, promise, reply, suggest, think, warn
<i>Elbaum (2001) Grammar in Context</i>	add, admit, announce, answer, claim, comment, complain, confess, declare, explain, reply	advise, ask, assure, convince, inform, notify, promise, remind, teach, warn	
<i>Fuchs&amp;Bonner (1995) Focus on Grammar</i>			advised, asked, invited, report
<i>Murphy (1989) Grammar in Use</i>	--	--	--
<i>Raimes (1998) Grammar Troublespots</i>		advise	ask, complain, reply, want to know
<i>Thewlis (2001) Grammar Dimensions</i>		order, ask, invite	

"Julie said Thomas was sick." Elbaum (2001) puts forth a "Rule of Sequence of Tenses," as follows, which the other textbooks conform to as well:

(know)	present	----->	simple past	(knew)
(am studying)	present progressive	----->	past progressive	(was studying)
(saw)	simple past	----->	past perfect	(had seen)
(was helping)	past progressive	----->	past perfect progressive	(had been helping)
(have taken)	present perfect	----->	past perfect	(had taken)
(had done)	past perfect	----->	past perfect	(had done)

Accordingly, the textbooks instruct students to shift modal verbs to a past form too by

adding a past-tense morphological ending to the modal. So *can* --> *could*; *may* (possibility) --> *might*; *may* (permission) --> *could*; *must* --> *had to*; *will* --> *would*. Textbooks also referred to some non-changing modal verbs such as *should*, *should have*, *could have*, and *must have*.

Textbooks gave various exceptions, or questionable/optional circumstances, to the tense backshifting process. The exceptions are listed below according to how frequently they were mentioned in the textbooks.

The Tense Backshifting process is ....

- *often not applied* when something is still present or true (Bland, 1996; Eastwood, 1999; Elbaum, 2001; Fuchs, 1985; Murphy, 1989; Thewlis, 2001)

- “He said the pizza is not here.”

- *usually not applied* when reporting soon after it was said (Bland, 1996; Elbaum, 2001; Fuchs, 1985; Thewlis, 2001)

- “I just talked to them and they said they are quite happy.”

- *not applied* when the main (reporting) verb is in the present tense (Bland, 1996; Elbaum, 2001; Fuchs, 1985)

- “Bob says he is sick”

- *often not applied* when stating a general truth (Elbaum, 2001; Fuchs, 1985; Thewlis, 2001)

- “She said Canada is larger than Portugal.”

- *often not applied* when future tenses are involved: when a future action has not happened yet, has already passed, or happened long ago (Bland, 1996; Elbaum, 2001; Thewlis, 2001)

- “She said she is going to start working on a Ph.D.”
- “ It was 1931 and he said he’s never going to trust the stock market again”

- *often not applied* to a past tense verb (Eastwood, 1999; Elbaum, 2001)

- “My grandmother told me that she was born before the war.”  
(from “I was born before the war.”)

- *often not applied* with hypothetical statements (Thewlis, 2001)

- “She asked me what is going to happen if I move to China.”

Another interesting specification about this Sequence of Tense rule and its exceptions occurs when one textbook (Murphy, 1989) informs students that they “*must* use a past tense when there is a difference between what was said and what is really true” (p. 92).

It does seem strange that the textbooks are so inconsistent in their presentation of exceptions. This seems to reflect choices made by the textbook writer(s) concerning the level of complexity to present to students. For instance, the fact that Raimes (1998) does not list any exceptions to the backshifting process coincides with the nature of that grammar reference; it is intended to be a concise grammar reference, and it would not want to over-explain complicated secondary details. We might also assume that the first exception listed above (no backshifting when something is still present or true) is the most widely accepted exception. Six of the seven textbooks refer to this exception. Subsequent exceptions, however, seem not to be so certain. Fewer textbooks present the other exceptions (no backshifting when stating a general truth, when using hypothetical constructions, etc.). This appears to indicate a degree of confusion about reported speech behavior. These textbooks communicate that it is not clear exactly how reported speech behaves.

*The reporting verb's tense* While textbooks give abundant advice about how to manipulate embedded clausal verbs (those verbs that occur within the reported speech), they do not seem to address how to decide what verb tense to use with the first verb—the reporting verb. Most textbooks seem to advocate a past tense reporting verb. They seem to indirectly imply an inherent logic of or their preference for a past tense reporting verb, as most examples use a reporting verb that is in past tense. Fuchs and Bonner (1995) even actually state directly that “we usually use the past tense of reporting verbs such as *say* or *tell* to report speech” (p. 271). They then go on to explain that they mention this because “whenever these verbs are in past tense, the tense of the verbs in indirect speech actually changes from the one the speaker actually used” (p. 271).

Even though most mention that the Sequence of Tense rule should not apply when a reporting verb is in present tense, some textbooks present all their examples with past tense reporting verbs. So, although textbooks refer to the fact that sometimes the reporting verb will

be in present tense (and then give instructions on what to do with the clausal verb in that situation), they do not give any reasons why or hints about when this might occur or when students should use a present tense reporting verb.

A few textbooks, in particular Eastwood's *Oxford Practice Grammar* (1999), present a more liberal view of reporting verb tenses. Eastwood prominently displays present tense reporting verbs in the first featured examples and then further throughout the later samples too. However, the text never addresses when or why or how to choose the tense of the reporting verb. Perhaps since many of the textbooks introduce reported speech as a re-characterization of what was said according to the reporter's point of view, textbooks assume that students will just make the appropriate decisions on their own regarding the context of the situation and the timeframe. One textbook briefly approached the issue of *when* (as in what context) present tense reporting verbs were often found, but did not give any indication as to *why* this occurs: "Note that in newspaper and magazine articles, reporting verbs are often in the present tense" (Fuchs & Bonner, 1995, p. 271). Another textbook (Bland, 1996) distinguished between a "present tense report" (using "says") and a "past tense report," (using "said") but again gave no indication of how one might choose between the two alternatives.

Overall, with all the exceptions and potential verb-tense-combination scenarios, textbook directions are quite complicated. To their credit, many textbooks made statements acknowledging that the verb tense choices are highly dependent on a given situation and the speakers' personal and time relationships to each other (Bland, 1996; Eastwood, 1999; Murphy, 1989). Textbooks also seemed fairly careful to hedge their instructions appropriately. Most managed to avoid characterizing their tense-shift instructions as absolute truths to always be followed, and instead incorporated a fair amount of "this usually happens"-type language. Fuchs & Bonner (1995), as an example, state that "verb tenses often [not always] change in indirect speech when the reporting verb is in past tense" (p. 282).

### **Rules explicitly given in textbooks**

The grammar textbooks mentioned an array of other reported speech instructions in addition to the reporting verb and tense shift advice. Some textbooks gave a few broad patterns, while others put forth a long list of detailed rules to follow. Most mentioned how

pronouns and place/time adverbials shift in reported speech and also addressed reported questions and reported imperatives, but beyond those similarities, any additional explicit rules varied widely. For instance, Raimes (1998) lists eight additional explicit rules including when to use quotation marks, how to punctuate, and how to reword uncompleted sentences in reported speech, while the only additional explicit rule that Thewlis (2001) gives is that *that* is used to introduce reported statements and is often omitted in informal contexts.

Perhaps Thewlis' directive is a sign that some of the textbooks acknowledge language varieties. As previously mentioned, Fuchs & Bonner (1995) pointed out that reported speech behaves differently in newspapers and magazines. Yet the textbooks do not refer specifically to varieties of language. Sometimes the textbooks allude to the possibility of register variation, but none focus any of their presentation on language varieties or register variation. This seems problematic, since language variety is so crucial to accurate language description. Are textbooks assuming and teaching that reported speech behaves the same way all the time—no matter what the situational context or register?

Table 4.2 illustrates textbooks' wide contrast in the additional explicit rules (those other than the reporting verb and tense shift rules) given about reported speech.

Table 4.2. Samples of Contrasting Additional Reported Speech Rules Given Explicitly in Textbooks

<b>Bland (1996)</b> <b>Intermediate Grammar</b>	- reported speech has a reporting verb in the main clause followed by a noun clause or an infinitive; - quoted statements change to a <i>that</i> -clause; - if reported verb is in present tense, the <i>that</i> -clause tense doesn't change; - "assure," "convince," "inform," "notify," "persuade," and "remind" must be followed by the listener; - a listener is optional with "answer" and "promise"; - pronouns and adverbs often must be changed to reflect the reporter's point of view; - when a <i>that</i> -clause follows the verbs "ask," "require," "insist," "demand," "suggest," "recommend," "advise," and "propose," the verb in the <i>that</i> -clause is always in the simple form even if those verbs are in the past
----- <b>Murphy (1989)</b> <b>Grammar in Use</b>	----- - especially with "tell" and "ask," use the infinitive for orders and requests; also with "said"

Another notable feature is that many books introduced the topic by saying that, in reported speech, the *general meaning* of what was said is more important than the exact words (Bland, 1996; Eastwood, 1999; Elbaum, 2001; Thewlis, 1997). It is not always clear exactly what the textbooks mean when they instruct students to convey “the general meaning.” It seems that the textbooks that mention this are trying to emphasize the circumstantial variation of each reported speech occurrence. They seem to be encouraging students to ultimately “do the right thing” and make an appropriate grammar choice based on common sense. Although this does not seem to be an explicit rule, it is interesting as a uniting piece of advice common among many different grammar textbooks.

In sum, the textbooks give students many directives—in addition to verb choice and verb tense—concerning how to use reported speech. Unfortunately, these numerous directives are often inconsistent from one textbook to the next. This results in significant confusion when describing reported speech behavior to students.

Although the additional directives cover many scattered topics, some consensus does exist among the multitude of advice.

*Common consensus textbook points:* the Sequence-of-Tenses Rules shifts reported speech to a past tense form following a past tense reporting verb; pronouns and time & place adverbials shift in reported speech to reflect real time; reported yes-no questions behave slightly differently and require special treatment. Additionally, textbooks do not indicate much register variation with reported speech.

### **Empirical Corpus Data Findings**

As evident from the previous textbook analysis, navigating a way through textbooks’ instructions of reported speech proves to be quite a complex task! This empirical corpus study should provide a more informed description of authentic reported speech. Though the corpus results will probably not simplify reported speech behavior, corpus results should give valuable frequency information. This frequency data can then be used to appropriately *prioritize* the presentation of reported speech to students. In fact, one underlying goal of the corpus study deals with analyzing naturally-occurring samples of the reported speech behaviors described in textbooks and then prioritizing the presentation of those behaviors according to

their frequencies in real life.

As mentioned earlier, the focus of the corpus study revolves around the frequency rates of reporting verbs and the combinations of tense shifts between a reporting verb and its clausal verb. Analyzing results in these two areas should enable a more informed, relevant, and hierarchically-appropriate presentation of authentic reported speech grammar to occur.

### **Frequency of reported speech**

To review, the corpus data itself, as described in the Methodology chapter, stemmed from two registers of American English: News and Conversation from the Longman Spoken & Written English corpus. This enormous amount of data was filtered through the computer program's search algorithm and then hand-edited to isolate reported speech occurrences. The results centered around the two main reported speech verbs, *say* and *tell* in News and Conversation. Other verbs occurred as reported speech as well, but these "other verbs files" were significantly smaller than the *say* and *tell* files, and so they will be discussed later in the "verbs occurrences" section of this chapter.

The main results consisted of reported speech occurrences which used *say* and *tell* in News and Conversation. As shown in Figure 4.1 below, *say* occurred as reported speech in News 4,205 times per million words and 1,011 times per million words in Conversation. *Tell* occurred as reported speech in News 340 times per million words and 317 times per million words in Conversation.

Figure 4.1 illustrates that reported speech occurs most often with some form of *say* (says, said, saying, etc.) in the News register. *Say* in Conversation is the next most frequent form of reported speech, while *tell* in both News and Conversation occurs at a much lower frequency. *Say* in the News register occurs as reported speech more frequently than all other reported speech forms combined. And across both registers, *say* is much more common than *tell* as reported speech. Both the News and Conversation registers use *tell* with reported speech at approximately the same rate; however, the News register uses *say* more than the Conversation register uses *say* to report speech.

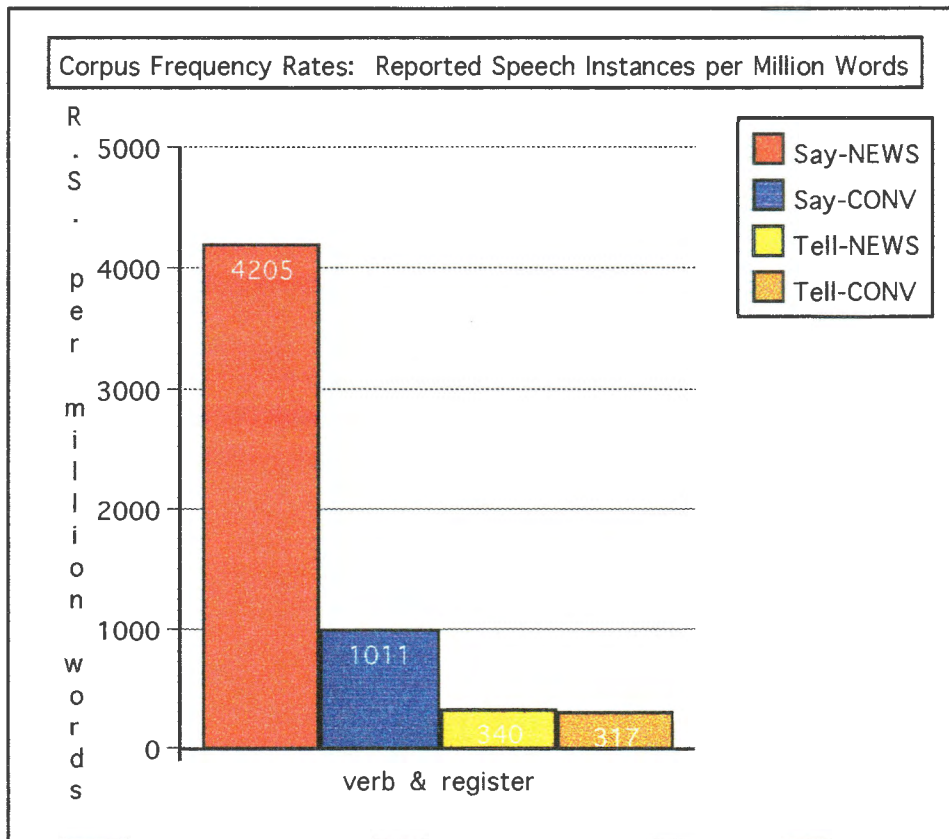


Figure 4.1. Overall Frequency of Reported Speech by Register

### Potential student problems with identifying reported speech

To arrive at the frequency results, true reported speech had to be distinguished from other similar grammatical constructions. This brings up the issue of students identifying reported speech. In order to determine these frequency rates, specific judgments had to be made about each encountered potential reported speech occurrence—specifically, whether or not the occurrence was reported speech. Not all of the decisions were clear-cut. Some occurrences were a bit tricky, and caused some second-guessing and confusion even for native-speaker linguists. This struggle is understandable given that all resulting instances being looked at had already been filtered through the computer program. Some occurrences, therefore, fit the standard reported speech patterns, but for some reason, were not actually reported speech. For example,



- *I sat there minding my own business and then she asked David later on she said that was your mother and he said ... (conv)*  
(confusion: seems to be direct speech)
- *If a doctor says that to me I say, well I feel ... (conv)*  
(confusion: hypothetical statement that hasn't happened yet -- in future, so can't be reported?)
- *I stopped there for, then I said I don't you know I'm leaving town. (conv)*  
(confusion: seems to be direct speech)
- *"Don't tell this member, whose wife has fought eight years against breast cancer, who's lying in a hospital now, that breast cancer research is unnecessary," said ... (news)*  
(confusion: imperative/command)

All of the above samples were considered problematic, but eventually judged to be not really reported speech. Various issues including direct speech, hypothetical statements, imperatives, made them difficult to judge at first glance. They caused uncertainty even at second and third glances, often. We can assume, then, that these sorts of instances could cause even more confusion for ESL students. Additionally, other similar structures that actually are reported speech, might be difficult to identify and use, such as these question-marked samples in the corpus:

- *Dan Shomron said Sunday it was originally planned to issue gas masks in December. (news)*  
(confusion: lack of explicit subject in embedded clause -- "dummy"/assumed "it" subject)
- *Why do they always say it's semifinalist instead of semi finalist? (conv)*  
(confusion: "always say" -- is actually report of what is always said, so OK)
- *and they said yeah, the baby has like Aids anti-, or Aids some kind of cells ... (conv)*  
(confusion: "yeah" makes it sound like direct speech, though it's probably reported speech, since the speaker in this version seems to be struggling to remember and rephrase details from the original utterance ["Aids anti- or Aids some kind of ..."])
- *well yeah, that's what Kathy told me that she moved you know. (conv)*  
(confusion: "what" confuses the underlying "Kathy told me that she moved" r.s.)
- *Every season there's at least one show the ad agencies and the network wise guys tell us we can't miss. (news)*  
(confusion: "tell" in relative clause; though OK because the statements have happened in the past and so are now being reported -- same as "wise guys have told us we can't miss")

These sorts of issues most likely represent significant areas of student comprehension and production difficulties with reported speech structures. Students will probably struggle

with these types of issues not only in classroom grammar homework, but also in everyday language usage when these sorts of structures occur. Students might communicate these types of messages incorrectly as well as misinterpret these types of grammatically encoded messages.

As you might recall from the previously discussed textbook analysis, the reported speech presentations in the textbooks did not seem to address these sorts of complicated issues. Their examples showed straightforward, clear-cut reported speech instances. A number of reported speech instances in the corpus results, however, were not so simple. It seems that the textbooks could use these questionable corpus samples to address problematic issues in reported speech for students, particularly advanced students.

### **Verb occurrences**

We can look at verb occurrences in two different ways: by proportion and by frequency. Both perspectives provide insights into the behavior of specific verbs, and considering both angles together gives a more complete picture of reported speech verbs. However, the concepts which these numerical differences represent can be somewhat confusing.

The *verb proportion* data tells us how much (what percentage of) reported speech occurred in the total filtered corpus file for each specific verb. That is, what proportion of our specific sample was reported speech. But because all of our specific samples were different sizes, their proportions cannot be compared accurately. Therefore, *verb frequency* data is needed to equalize all the varying sample sizes and show us a more equal comparison. The verb frequency data is given according to “occurrences per million words.” Since the frequency data has been normed per million words, this data gives more of an overall big picture within the grand scheme of the entire corpus—or at least the total scope of each particular register within the entire corpus. In contrast to the limited proportional data, we are able to apply this normed frequency data more liberally, since it is truly representative. For instance, using the same verb, *claim*—in the Conversation register, the normed frequency data shows that reported speech instances using the verb *claim* occur in Conversation about four times per million words. We can, therefore, fairly safely say that *claim* will likely be used as reported speech in American Conversation approximately four times per million words.

Although *claim* has a fairly significant proportional rate above 50% (53%) in Conversation, the number of total instances is quite small (15 total pattern-fitting-instances compared with *say*-Conversation's 4,222 pattern-fitting-instances). The normed frequency rates are able to account for such drastic variables, and therefore give a more balanced perspective in terms of the entire corpus.

The relationship between proportion and frequency rates may appear to be indicating that the normed frequency rates carry more importance. While the normed frequency rates are "superior" in many ways, we can still learn from the proportional rates as well. In fact, the normed frequency rates seem to speak for the big, pervasive reported speech verbs as their constituency, whereas the proportional rates seem to give voice to the smaller, less common reported speech verbs. Even though some high-proportion reported speech verbs may have small frequency rates, when those infrequent verbs *are* used in the pattern-fitting way, some of them are extremely likely to be reported speech even though they may not occur all that often in the grand scheme of the whole corpus. For instance, *persuade* does not occur all that often (14 occurrences in filtered file) in News compared with *tell* (1,570 occurrences in filtered file), but when *persuade* does actually occur pattern-filtered in News, it is more likely to be reported speech (79%) than *tell* is (70.8%).

At this point, we briefly examine the data results using both perspectives. Table 4.3 shows both frequency and proportional data for the two main verbs (*say* and *tell*) along with some smaller, "other" verbs.

Table 4.3 shows that *say* and *tell* emerged as the most common reporting verbs, by far. They are used much more than any other verb to report speech in both News and Conversation registers. Since *say* (4205 and 1011 occurrences per million words) is more frequent than *tell* (340 and 317 occurrences per million words), quantitative findings point to *say* as the most frequent reporting verb throughout the entire corpus sample.

Also shown in Table 4.3, the "other verbs" exhibit a wide variety of frequencies and proportions. The "other verbs" listed in the above table are those with high proportion or frequency results. The additional "other verbs" are not listed in the above table because their numerical results were below a 50% *proportional* rate and their *frequency* rates were below 20 occurrences per million. (Details about these additional non-listed "other verbs," as well as

Table 4.3. Noteworthy Verb Findings from each register  
(verbs that exhibited high proportional *or* high frequency rates)

<u>NEWS</u>		
	<b>Proportions %</b> (Percentage of r.s. among pattern-fitting occurrences)	<b>Frequencies</b> (r.s. occurrences/million words)
<b>Say</b>	76.5%	4205
<b>Tell</b>	70.8%	340
Others		
add	31%	33
admit	40%	23
<b>agree</b>	67%	39
announce	41%	78
<b>claim</b>	55%	51
<b>indicate</b>	66%	60
report	18%	78
reply	67%	5
swear	57%	2
ask	48%	76
assure	62%	9
convince	50%	4
inform	54%	5
notify	58%	5
persuade	79%	3
<b>warn</b>	80%	34
 <u>CONVERSATION</u>		
	<b>Proportions %</b> (Percentage of r.s. among pattern-fitting occurrences)	<b>Frequencies</b> (r.s. occurrences/million words)
<b>Say</b>	49.6%	1011
<b>Tell</b>	44.4%	317
Others		
claim	53%	4
mention	24%	8
swear	57%	8
advise	100%	0.5
ask	27%	61
assure	75%	1
warn	67%	1

complete results details, can be found in Appendix C.) Yet even the highest frequency “other verbs” do not come close to approaching the frequency rates of *tell* or *say* in either the News or Conversation register. Among the “other verbs,” *report* and *announce* have the highest reported speech frequency rates at 78 occurrences per million words, but that frequency rate is still more than four times below *tell*’s lowest frequency rate (317 occ./mill. in Conv.) and neither *report* nor *announce* are at all frequent in the Conversation register—only in the News register. So while *say* and *tell* certainly are not the only verbs used with reported speech, these results reinforce the idea that *say* and *tell* most definitely account for the largest amount of reported speech verbs.

**Register variation with reporting verbs** Perhaps most significant in the verb findings is the extent to which verb occurrences vary according to register. Particular verbs behave quite differently depending on whether they are used in the context of conversation or news. *Say* and *tell* exhibit some register variation, with *say* varying more than *tell*’s steady behavior. *Say*’s frequency rate drops drastically from News to Conversation (4205/mill. to 1011/mill.) compared to *tell*’s relative stability (340/mill. to 317/mill.) between the two registers. This seems to indicate that people use *tell* approximately as often in News as in Conversation, but people use *say* four times more often in Conversation than they do in News. Once again, this may point to a more controlled type of language use in the News register (a style of language which would want to avoid overuse of one word, “say,” or that would want to use more specific, particular vocabulary such as “she claimed, he notified, she informed, they agreed ...”) and a more casual, repetitive, informal, simplistic type of language use in Conversation registers.

Interesting discrepancies exist among the smaller “other verbs.” Specifically, some of the less frequent “other verbs” such as *add*, *report*, *warn*, etc. fluctuate dramatically from one register to another. For example, *warn* in the News register had a frequency of 34/mill., and the highest proportional rate in the News register at 80%. In conversation, however, *warn*’s frequency drops to just 1/mill. with a proportional rate of 67%. A more dramatic example can be seen with the verb *add*. Within Conversation, the verb *add* occurred 95 times in potential-reported-speech conditions, but none of the instances were actually reported speech. So while

*add* in Conversation had a 0% reported speech proportion rate, in News *add* as reported speech had a 31% reported speech proportion rate, occurring 108 times as positive reported speech out of a potential 354 instances. This means that *add* in Conversation occurs zero times per million words, while *add* in News occurs 33 times per million words.

Therefore, it seems impractical to make generalizations concerning how common “other non-say-tell verbs” are in reported speech; rather, any attempt to give this sort of information can only happen accurately in the context of specific registers. So it would be legitimate to say that *reply* ties with *announce* as the most common “other reporting verb” in News. But how could we possibly make a blanket statement identifying *reply* as one the most common reporting verbs overall? After all, *reply* is among the least frequent in the Conversation register—it did not occur at all in the Conversation samples (or subcorpus) (0% proportion and 0/mill. frequency)!

Specifics and implications of register variations will be discussed in more depth in a further section of this chapter, but it is important to keep in mind such issues when discussing general frequency rates and file proportions of reporting verbs. In terms of the general frequencies of reporting verbs, *say* and *tell* lead by a wide margin, as mentioned previously. Only a few other verbs manage to appear in both News and Conversation registers as noteworthy, including *ask* with a high frequency rate in both registers, and *claim*, *warn*, *swear*, and *assure* with high proportional rates in both registers. Numerous other verbs stand out in one or two categories in one of the registers, but not in both registers. Frequency rates probably give the most reliable indication of “other verb” importance; in that case, only five “other verbs” qualify as important—with frequencies of at least 60 reported speech occurrences per million words. These include *announce*, *report*, *ask*, and *indicate* in the News register and *ask* in the Conversation register.

### **Tense combinations**

The clearest way to present verb tense combinations, for these purposes, is to divide the possibilities into four broad, simplified categories: *Past-Past*, *Past-Present*, *Present-Past*, *Present-Present*. As discussed in the Methodology chapter, these four categories reflect how the reporting verb tense and the embedded clause tense combine, on a general level, in a

reported speech occurrence. For example, “*Laura said she had seen Joel*” would be classified as following a general “Past-Past” pattern and a detailed “Past-->past perfect” pattern (See Appendix B for a reminder of these tense breakdowns). Both levels of classification highlight particular information and give unique insights into various verb tense combination issues.

With these hierarchical levels of classification clarified again, we can proceed with presenting the data and analyzing the results using these two systems. Table 4.4 shows the simplified breakdown of tense combinations per register.

Table 4.4. Verb Tense Combinations per Register

		<b>PAST</b>		<b>PRESENT</b>	
		<b>Past-Past</b>	<b>Past-Present</b>	<b>Present-Past</b>	<b>Present-Present</b>
<b>Say-NEWS</b>					
Total:	413	245	94	20	54
Percentages:		<b>59.32%</b>	<b>22.76%</b>	<b>4.84%</b>	<b>13.08%</b>
<b>Say-CONV</b>					
Total:	100	51	31	4	14
Percentages:		<b>51.00%</b>	<b>31.00%</b>	<b>4.00%</b>	<b>14.00%</b>
<b>Tell-NEWS</b>					
Total:	100	67	20	5	8
Percentages:		<b>67.00%</b>	<b>20.00%</b>	<b>5.00%</b>	<b>8.00%</b>
<b>Tell-CONV</b>					
Total:	101	64	32	3	2
Percentages:		<b>63.37%</b>	<b>31.68%</b>	<b>2.97%</b>	<b>1.98%</b>

Figure 4.2, which presents the same information in a more visual way, illustrates that, in general, Past-Past is the most frequent tense combination, but never accounts for more than 67% of the tense combinations in any register. Past-Present is the next most frequent, followed by Present-Present and Present-Past, respectively (except with *tell* in the Conversation register, where Present-Past is slightly more common than Present-Present; it should be noted, though, that the percentage numbers for those two particular tense combinations are extremely low with *tell* in the Conversation register. Only five of the total 100 occurrences in that register were anything other than Past-Past and Past-Present).



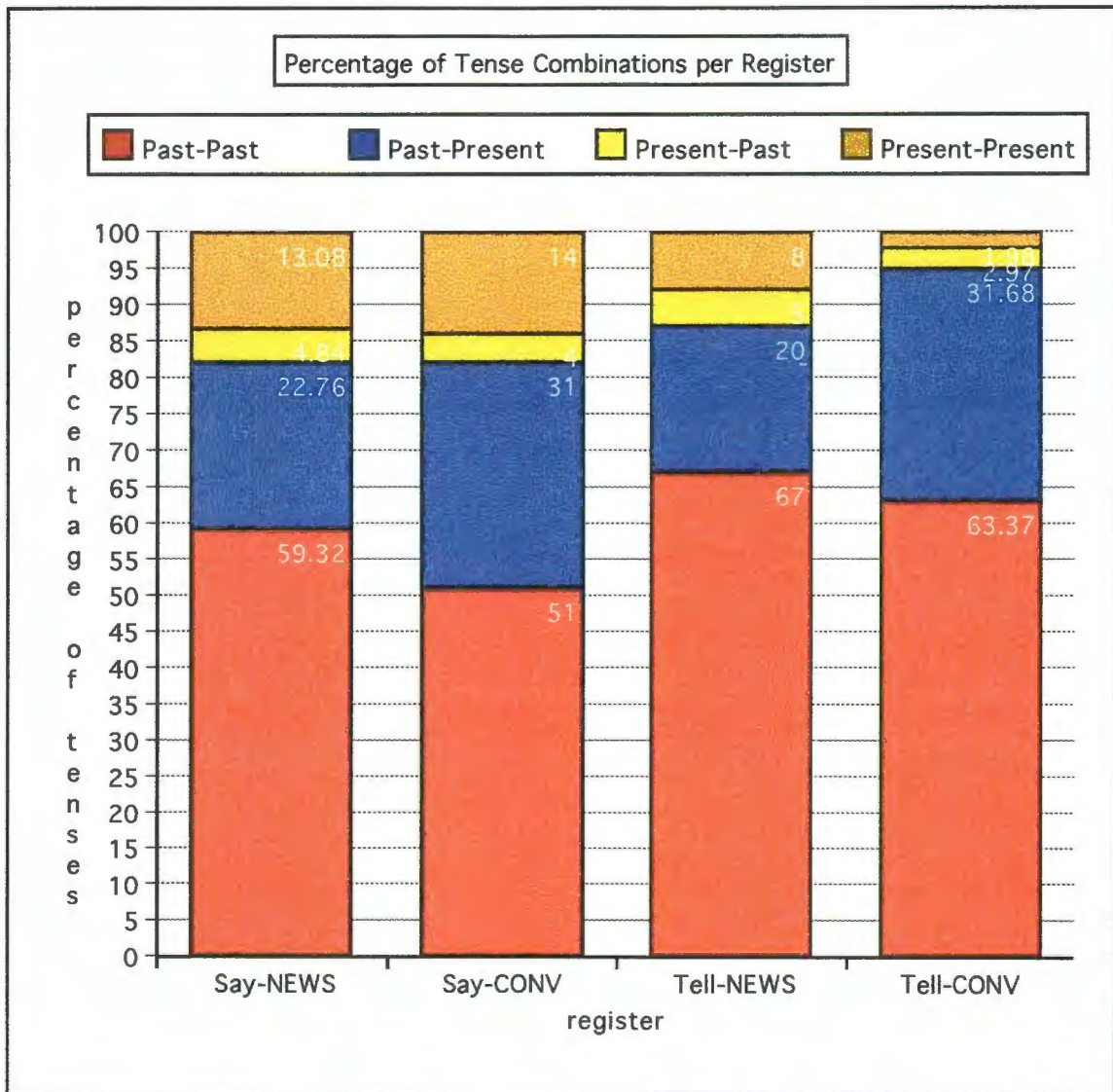


Figure 4.2. Proportional Tense Combinations

At this most basic level, these results show that a Past-Past tense combination, which indicates that the traditional textbook advice has been followed,<sup>3</sup> is the most common tense

<sup>3</sup> The Past-Past tense combination represents the traditional textbook advice for the following reasons: 1) It would be impossible for Past-Present or Present-Present to have been backshifted to past tense (since their embedded reported speech clauses are clearly in present tense!). 2) Textbooks usually specify that the main clause reporting verb should be in past tense, so a Present-Past combination would apparently contradict the customary textbook advice. So, given (1) and (2), we can deduce that the verbs have been backshifted in these Past-Past tense combinations, and, therefore, represent the traditional textbook instructions. 3) Even in the case of a non-backshifted verb (as with *John*: "I was sick." ---> *John said he was sick*), this conforms to some textbooks' "exception" instructions (Eastwood, 1999; Elbaum, 2001) which direct students to not apply backshifting in such cases. And if it did backshift to *John said he had been sick*, it would still categorize as Past-Past.



combination overall—for both *say* and *tell* in both News and Conversation registers. We must note, however, that some Past-Past combinations may not have been backshifted if the original utterance’s clausal verb was in past tense, as in *John: I was sick*. The backshifted report would be “*John said he had been sick*.” Yet it seems quite plausible that a speaker might instead report the statement as “*John said he was sick*.” However, we have no way of knowing if the original utterance corresponding to the report “*John said he was sick*” was *John: I am sick* or *John: I was sick*.

But the other tense combinations play quite significant roles as well! Other, non-customary (according to textbook directions) tense combinations occur in almost half of all *say* reported speech instances. In other words, for instance, with the verb *say*, the traditional backshifted Past-Past accounts for just slightly more than 50% in both News and Conversation registers. So, although the traditional textbook directions regarding tense combinations do hold true in the majority of cases, they are only followed a little more than half the time (since the majority margin is not very big). Other non-traditional-textbook tense combinations occur in 40.68% of the Say-NEWS instances, 49% of the Say-CONV instances, 33% of the Tell-NEWS instances, and 37% of the Tell-CONV instances.

These results, though similar in their general proportions, show distinct patterns emerging between *say* and *tell* behaviors. While neither *say* nor *tell* behaves overwhelmingly traditionally, conforming to textbook backshifting conventions, *tell* does tend to do so more often than *say*. *Tell*, then, becomes identified as a verb with more traditional tense combination behavior (67% and 63% Past-Past) as compared with *say* (59% and 51% Past-Past).

Regarding the four basic verb tense combinations, some specific structures within the broad categories emerged as common reported speech forms. Regular, or “simple-simple,” combinations never accounted for much more than 70% in the broad four categories. Furthermore, modal usage seems to be a factor influencing verb tense combinations in reported speech. Some of the significant factors are highlighted below in Table 4.5.

Within the general Past-Past category, the regular past-->past tense combination was the most common detailed tense combination (such as “*Stacy said he was really nasty*” [conv]). Two other detail-level combinations were relatively common too: 1) a simple past tense verb to a past-Modal + a present verb, *past--> M-past+present* (such as *She told them today she*

Table 4.5. Notable Highlights of Detailed Verb Tenses

	Say-NEWS	Say-CONV	Tell-NEWS	Tell-CONV
<b>PAST-PAST</b>				
Regular past-->past	56%	51%	57%	59%
past-->M-past+present	18%	14%	16%	25%
past-->past perfect	7%	16%	10%	2%
other	19%	19%	17%	14%
<b>PAST-PRESENT</b>				
Regular past-->past	56%	68%	65%	63%
past-->M-present+present	18%	10%	10%	9%
past-->present perfect	10%	10%	15%	6%
other	16%	12%	10%	22%
<b>PRESENT-PAST</b>				
Regular present-->past	80%	25%	60%	33%
other	20%	75%	40%	67%
<b>PRESENT--&gt;PRESENT</b>				
Regular present-->present	59%	57%	75%	50%
other	41%	43%	25%	50%

*couldn't handle the stress on her first day* [conv]) and 2) a simple past tense verb to a past perfect verb, *past--> past perfect* (such as “*He said he hadn't washed it out yet though* [conv]).

Similarly, in the Past-Present category, besides simple past to simple present (such as “*He said a million Americans are in prison*” [news]), the two other most common categories involved a modal and a perfect tense: 1) a simple past verb to a present modal + a present verb (past--> M-present+present) such as “*Thomas Fallon said he will not budge*” (news) and 2) simple past to present perfect (past-->present perfect) such as “*They said they've got a space*” (news).

Present-Past categories were all quite small, and so no detailed tense combinations emerged clearly as in the previous two tenses. However, a consistent pattern appeared in News with both *say* and *tell*--- simple present to simple past were the most common tense combination (such as “*People in Comaneci's entourage say the two broke up en route back from Japan*” [news]). Both *say* and *tell* in the Conversation register had other, “non-simple” detailed tense combinations as the most common forms, though.

Present-Present did not show any patterned detailed verb combinations besides the regular present-->present forms in either of the registers. Simple present to simple present

accounted for most of the Present-Present occurrences, such as “*The National Indian Office says it fears the Yanomami tribe*” (news).

Because the frequency rates for the other “non-say-tell” verbs were so comparatively low, I did not examine those verbs’ tense combinations in detail. Future research involving more occurrences of these verbs would be useful, though, to determine if their tense patterns follow the same pattern as the *say* and *tell* verb patterns.

**Modals in the tense combinations** As noted previously in Table 4.5, modal verbs emerged as distinctive in the results. Although this present study did not analyze modal usage in depth, future research endeavors could investigate modal usage further. Table 4.6 shows the prevalence of modals with *say* and *tell* in both registers. These results show that modals are a steady component (slightly above or approaching 20 percent in all cases) of reported speech verb combinations—at least with *say* and *tell* in the News and Conversation registers.

Table 4.6. Modal Usage

	Total # Reported Speech	Occurrences with Modals	Percentage of Modal Usage
Say-NEWS	413	85	20.58%
Say-CONV	100	19	19.00%
Tell-NEWS	100	16	16.00%
Tell-CONV	101	21	20.79%

### Register variations

As stated earlier, the most significant finding of these corpus data results come from looking at register variations. Analyzing the most common reported speech forms across registers reveals that a single description of reported speech verb behavior is misleading. We need descriptions in terms of specific registers and specific verbs. Figure 4.3 (below) illustrates the register variation.

On a holistic level, reported speech, in general, happens more in News than in Conversation. We know this by looking at the frequency counts in each register. Not only does News use a wider variety of reporting verbs, but reporting verbs occur more frequently in

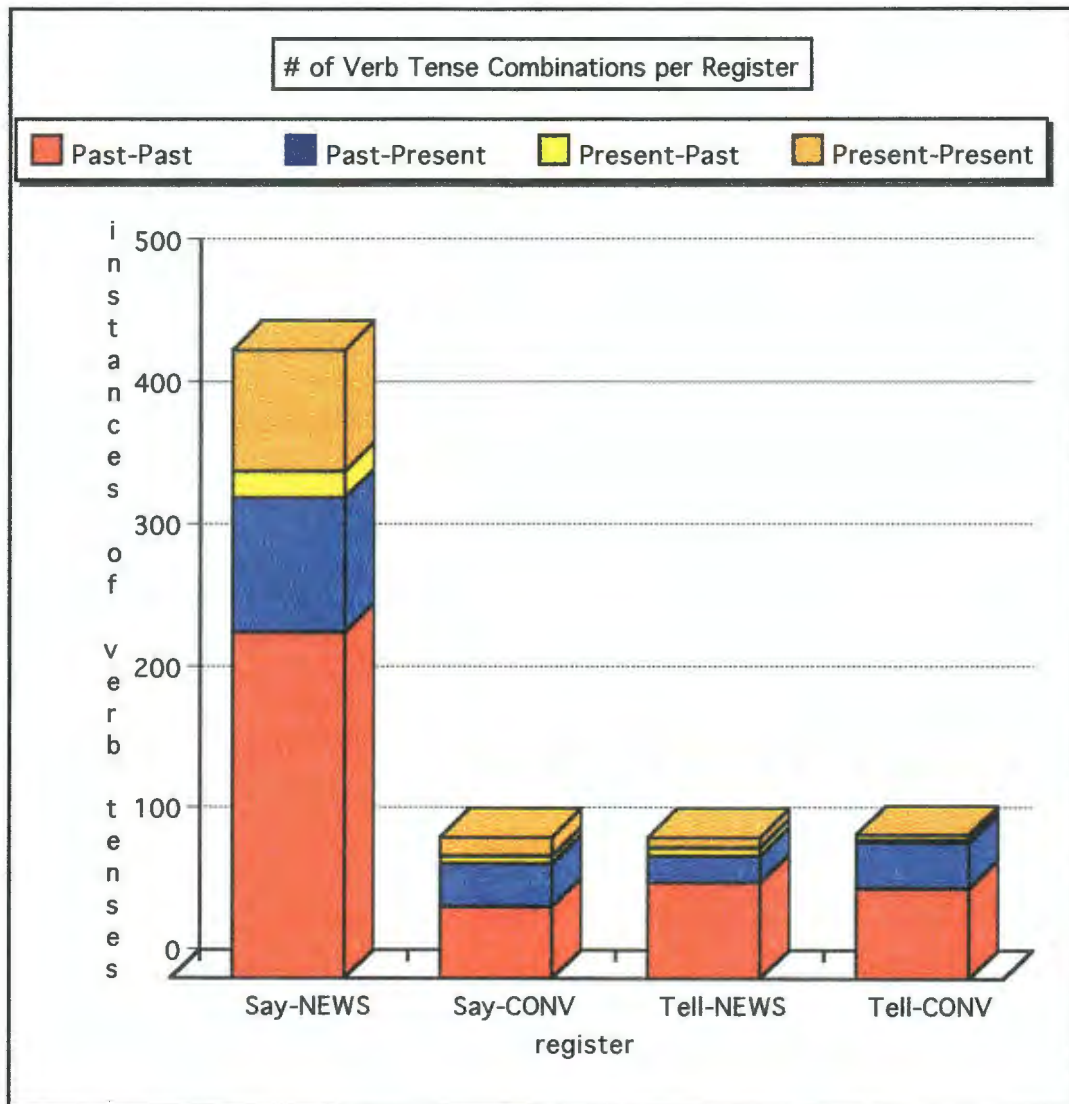


Figure 4.3. Basic-Level Verb Tense Combinations per Register

News. In fact, there are many more reported speech occurrences of the verb *say* in News than there are total reported speech occurrences in all of Conversation. Reported speech in News is approximately 3.5 times more common than reported speech in Conversation (~4,500 occurrences per million words in News and ~1,300 occurrences per million words in Conversation; see Figure 4.1). We could safely say that, typically, seven or eight of every ten reported speech occurrences will occur in News, while two to three occurrences will occur in Conversational contexts. This finding has fairly important practical implications which I will discuss in further depth later in this chapter.

Register variation is also important when we consider the choice of reporting verb. Both the News and Conversation registers utilize *say* and *tell* as the most common reporting verbs. However, people use the less frequent “other verbs” quite differently depending on which register the language occurs in. As described in the previous “verb occurrences” subsection, most of the “other reporting verbs” occur and behave differently in News and in Conversation. For one thing, people tend not to use the “other verbs” very frequently at all in Conversation. Only one of the noteworthy Conversation “other verbs” could be referred to as frequent: *ask* with 61/mill. frequency rate. So basically, we normally hear a much smaller variation of reporting verb vocabulary choice in Conversation than in News. Other verbs in News, although not nearly as frequent as *say* or *tell*, did at least exhibit relatively high frequencies, indicating that reported speech in News contexts tends to use a wider variety of reporting verb vocabulary choices (such as *agree*, *claim*, *indicate*, *warn*).

Some typical/representative corpus samples below illustrate and reinforce these register trends:

#### News samples:

“The International Energy Agency advised Western nations to brace for a possible oil shortage.”

“Not everyone agrees the system is desegregated .”

“Croatian leaders claim the unrest has been orchestrated by the Communist-ruled republic of Serbia to destabilize democracy in Croatia .”

“The Bureau of Indian Affairs indicated the 25 percent set aside for non-members of the tribes was almost ready for distribution.”

“The official Iraqi news agency reported that Hussein released the nine Frenchmen as part of a ‘humanitarian gesture.’”

#### Conversation samples

“I said I’m gonna kill them.”

“They said that Reagan had a plan.”

“He said you should have it checked every day.”

“He told me that he was really in graduate school ...”

“... Kathy told me that she moved you know.”

“Eric says we’ll be able to fix that.”

As for register variation in terms of verb tense combinations, both *say* and *tell* patterns use a higher percentage of Past-Past forms in News than in Conversation. They also use a

higher percentage of Past-Present forms in Conversation than in News. Another characterization would be that both *say* and *tell*, in terms of the four broad tense categories, tend to behave a little more radically, or non-conventionally, in Conversation than in News. This means that more traditional textbook backshifting seems to occur in the News register.

The only definitive register variation pattern among the detailed verb tenses occurs with the Present-Past category, where the News registers strongly tend to use the “regular” simple verb forms while the Conversation registers strongly use the non-regular forms (such as present-perfect-->past, present-->M-past+present, etc.). For example:

#### News:

simple present ---> simple past:

*“Wags from the east tell us Whoopi Goldberg ... almost outshone the nine hopefuls ...” (news)*

simple present ---> simple past

*“Actress Helen Hayes says Irving Berlin saved her marriage plans ...” (news)*

#### Conversation:

present ---> M-past + present perfect

*“He says it should have closed by now.” (conv)*

present ---> M-past + present perfect

*“...and they say they wouldn't have done anything if they caught him?...” (conv)*

The numbers regarding these patterns are quite small, though, and so may well be insignificant. No definitive results can be claimed until future work is done to study more occurrences.

### **Conditions influencing particular reported speech behaviors & patterns**

A range of situational variables surround each reported speech occurrence. Variables can include the verb tenses of the previous sentences, the speaker's tone of voice, the prior content or context of a conversation or reading passage, the shared experience between a speaker and listener, gestures, factors from the external situational environment (such as the weather or a building or piece of art or background knowledge). Realistically, not all of the



variables can be known from a printed corpus sample, and so we cannot expect to analyze these unknown variables. We can, however, search for, identify and then analyze variables that might be visible in a printed corpus format. Examining these variables can give insights into the choices made with verb tense combination and reporting verb selection.

***Influential Present-Past conditions*** An examination of the 30 Present-Past reported speech occurrences reveals that, once again, register variation surfaces as an issue. Consider some representative samples below:

TELL:

News

- ... wags from the east tell us Whoopi Goldberg, in high heels and sewed into a short, red-sequined dress by Nolan Miller, almost outshone the nine hopefuls ...

SAY:

Conversation

- ... they say they wouldn't have done anything if they caught him? ....
- ... He says it should have closed by now. ...
- ... It says that he um, he did a few operas and he gave up on operas ...

News

- ... California news black film actor says Police Held A Gun To His Head ...
- ... McQueen says PCBs from leaking transformers could leach downhill ...
- ... actress Helen Hayes says Irving Berlin saved her marriage plans ...

While both *tell* and *say* exhibit the Present-Past verb tense combinations, they seem to do so under different conditions. The rare *tell* examples tend to be more specialized uses of the verb *tell*—either colloquial uses or particular figures of speech. The Present-Past *tell* sample in the News register is certainly reported speech, but falls under the category of slightly unconventional: “wags from the east tell us Whoopi Goldberg...almost outshone ...hopefuls.” The phrasing of this utterance seems strange—as if it is an alternative way of communicating that “wags from the east have told or are always telling us that Whoopi Goldberg almost outshone hopefuls.” The *tell* usage in this statement conveys that the wags from the east still hold a certain opinion, or *claim or maintain* that Goldberg outshone hopefuls. Interestingly, this method of stating the information also tends to occur in News with *say*.

So with this last *tell* Present-Past sample being the only definitive reported speech

sample, it seems that we can conclude that a Present-Past tense combination with *tell* is extremely rare. Present-Past with *tell* occurs as a unique alternative form. With *say*, however, the Present-Past combination is more prevalent. Additionally, the Present-Past constructions with *say* seem to follow some logical patterns.

Within *say*'s Conversation register, the first two samples seem to convey hypothetical actions that were not completed, and so are non-concrete, as seen in: "they say they wouldn't have done anything" and "he says it should have closed by now." Additionally, both occurrences of *say* in these samples represent a sort of action also akin to the word "claim," so "he claims they wouldn't have done anything" or "he claims it should have closed by now" could also work. This sort of substitution seems to be acceptable because the speakers are using the word *say* to express an opinion held by the person they are reporting about. This occurs repeatedly in the *say*'s News samples as well: "California news black film actor says (or *claims* or *maintains*) police held a gun to his head" ; "the senate judiciary committee says (or *claims* or *maintains*) the U.S. led all countries" ; "McQueen says (or *claims* or *maintains*) PCBs from leaking transformers could leach downhill" ; Killea says (or *claims* or *maintains*) the measure was not intended to ban sales" ; "Helen Hayes says (or *claims* or *maintains*) Irving Berlin saved her marriage; "they say (or *claim* or *maintains*) New England should not experience the flurry of bank closings." These, and most *say* Present-Past samples in News, express the message that the parties claim, or consistently maintain, whatever message is being reported. This sort of external condition seems to cause, or at least permit, the use of a Present-Past verb tense combination, especially with *say* and perhaps sometimes with *tell*. Yet the textbooks from the previous textbook analysis did not mention anything regarding this reported speech condition or verb tense combination.

Some other particular factors in some specific samples emerge as well. For instance, "it says that he did a few operas" seems logical because the person appears to be currently reading an explanation at the moment and reporting what she has just read. What she has just read is what the item currently says and what it will continue to say. So perhaps this falls under the category of something that is still present or true, as the textbooks mention. However, textbooks present that condition as a reason to leave the embedded clausal verb in present tense (it said that he does a few operas), not to state the reporting verb in present tense



and backshift the clausal verb (it says he did a few operas). We must consider the factor of accurate and truthful timing as well, then. Since the discussion centers around Beethoven, obviously, any of Beethoven's musical compositions were written by him in the past, and so perhaps we are obligated to always say Beethoven *did* operas, since to say Beethoven *does* operas would imply that he is alive and currently doing operas, which is impossible. It seems that both of these justifications apply, and they both contribute to the unique necessity for such a Present-Past verb tense combination. In this situation, to use reported speech in the customary textbook way would amount to a fundamental difference in meaning between statement and reality.

It seems that Present-Past is a sort of customary way of conveying particular types of information—information that the reporter believes to be truthful and applicable to his present timeframe. (“Wags from the east tell us” = “Wags from the east have told us and still continue to tell us that ...”; “McQueen says PCBs could leach downhill” = “McQueen has said and continues to claim that PCBs could leach downhill.”) This may be worth noting or mentioning in textbooks—or at least familiarizing students with through exposure in examples. This pattern is especially relevant to the textbooks since many of them mentioned an element of “present truth” or “general truth” in their exceptions to tense backshifting. Such exceptions instructed students not to backshift, but did not mention that the reporting verb should be in present tense. But this corpus data shows that the element of present truth influences the reporting verb as well as the embedded clausal verb. Given these findings, textbooks should consider addressing both verb tenses in their presentations to students.

With proper resources, a similar sort of analysis could be done to investigate conditions surrounding modal usage, reporting verb choice, or other notable findings. However, given the scope of this study, the investigation into Present-Past serves as a sort of sampling of these types of efforts. A future study with more occurrences of the less common reporting verbs might also give insight into these matters.

## Discussion

Although the quantitative results of the study may be interesting on their own, the study itself gains more legitimacy when interpreted and applied to real life contexts. This discussion section will summarize and clarify the study's results so that the following chapter can discuss what the data and results mean and why these results might be important to others. Specifically, this discussion revolves around the initial research questions. Conclusions in chapter five then move on to more general implications from the research questions.

### Answers to Research Questions

Since most of the confusion about reported speech seems to deal with issues of reporting verbs and verb tense combinations, the research questions addressed those topics. This corpus investigation also provided an ideal opportunity to investigate register variation in another research question. These research questions are answered in terms of both textbook data and corpus data, thereby enabling us to assess textbook authenticity.

#### **Research question #1: What does an empirical corpus study reveal about the usage patterns of English reported speech?**

This section succinctly summarizes the corpus data results with regards to verb tenses, reporting verbs, and register variation. In brief, we discover that reported speech uses all four verb tense combinations, that *say*, *tell*, and a few other verbs, respectively, are the most frequent reporting verbs, and that reported speech behavior varies between the News and Conversation registers.

*(1.a) Which tenses are used in reported speech?* Empirical corpus data reveals that practically all types of tense combinations occur in reported speech. Of the four broad, basic categories, Past-Past is the most common overall, followed by Past-Present, Present-Present, and Present-Past. Even at its highest rate, the most common category, Past-Past, does not occur more than 67% of the time. This means that the classic backshifting of tenses (*Li said the government was intensifying efforts to complete a new plan -- news*) does

not seem to occur in any more than ~65 to ~70% of reported speech instances. Past-Present (*Aspen said Thursday that he plans to hold hearings on the disintegration of the Soviet army -- news*) and Present-Present (*News reports say many young women also take their lives to escape harassment ... -- news*) combinations occur at moderate rates, but a Present-Past combination (*Authorities say O'malley slashed his throat and shot him in the face because ... -- news*) is rather rare in reported speech (although it does exist).

Within the Past-Past and Past-Present categories, certain specific details of tense combinations exhibit patterned trends. A simple past tense reporting verb combined with an embedded clause modal verb + present tense verb seems to be a common form in both Past-Past and Past-Present. Also, a simple past tense reporting verb combined with a perfect tense embedded verb is another common form in both Past-Past and Past-Present.

**(1.b) Which reporting verbs are used in reported speech?** Verb frequency counts and proportional occurrence rates give the best answer to this question. *Say* occurs as the most common reporting verb, followed by *tell* which is also quite common, though not as frequent as *say*. Other reporting verbs do exist, but are much less common than *say* and *tell*. Moreover, specific “other verbs” vary highly according to register. The “noteworthy” other verbs found include: *agree, claim, indicate, warn*. Some less noteworthy, but somewhat noticeable other verbs include: *add, admit, announce, report, reply, swear, ask, assure, convince, inform, notify, persuade, mention, advise*. A few more “other verbs” appeared as reported speech in the corpus samples, but were very rare. These include: *complain, comment, declare, explain, promise, remind, teach*.

Of course, all of these “other verbs” came from textbook assertions. Therefore, the corpus study only searched for and found *say, tell*, and other textbook verbs. But a major portion of this study aimed to check the natural usage patterns of textbook reported speech verbs, so using textbook verbs makes sense. However, in future empirical reported speech studies it would be beneficial to look for other possible reporting verbs besides the 45 (43 + *say + tell*) textbook verbs used in this study.

(1.c) *How do reporting verbs and tenses vary across registers?* In order to answer this question we must examine the most common forms of reported speech in each register. The previous discussion about these matters showed that, in the corpus data results, reported speech varies in more than one way across registers.

To begin, reported speech seems to occur more often in the News register than in the Conversation register. Reported speech in News is ~3.4 times more common than reported speech in Conversation. Almost 80% of the corpus' reported speech instances (normed frequency instances) occurred in News. Clearly, this finding seems to reflect the inherent nature of News registers and the fact that News probably relies on reporting verbs more often than most other types of discourse.

In terms of reporting verb choice, *say* is the most common reporting verb in both News and Conversation, followed by *tell* as the second most common in both registers. Reported speech employs many other reporting verbs as well, but not even close to the same extent (in terms of frequency) as it uses *say* and *tell*. Only a few of those “other verbs” emerge as significant enough to selectively identify, and their significance depends on the register context besides. For example, *agree*, *claim*, *indicate*, and *warn* could be considered somewhat frequent (more than 30/million) as “other verbs” in News, while *ask* is the only frequent “other verb” (61/million) in Conversation.

In terms of verb tense combination choices, both *say* and *tell* behave more traditionally in News than in Conversation. Both verbs employ a higher proportion of Past-Past (“traditional”) tense combinations in News than Conversation, and both utilize a higher proportion of Past-Present (“semi-traditional”) tense combinations in Conversation than in News. Such Past-Past combinations (past reporting verb & past embedded clausal verb) indicate the traditional textbook-advocated “backshifting” of the embedded clause verb to a past tense form; Past-Past combinations also indicate that the main reporting verb has been stated in the past tense to reflect the appropriate time relationship:

- “traditional” past report backshifted: *He said he was sick*
  - as opposed to a “semi-traditional” past report not backshifted: *He said he is sick*
  - or a “non-traditional” present report not backshifted: *He says he is sick*
  - or a “radical” present report that may have been backshifted: *He says he was/had been sick*
- Say* and *tell* demonstrate more unique behavior in Conversation also by using more non-simple

verb forms than in News. In other words, in the Past-Past broad category, for instance, more “simple past --> simple past” combinations occur in News than in Conversation. Conversation tends to have a little more tense detail diversity, using more complicated verb tenses like past progressive, past perfect, past perfect passive, past modals.

## **Research question #2: How do EFL/ESL textbooks present reported speech?**

This section briefly summarizes the textbook analysis results with regards to reporting verbs, verb tenses, and real language authenticity. In brief, we discover that textbooks are neither entirely authentic nor unauthentic in their reported speech presentations.

### ***(2.a) Which reporting verbs are used in textbook reported speech?***

Textbooks present *say* and *tell* as the most common reporting verbs and also focus on those two verbs most often. Many textbooks also list additional verbs. Textbooks that do list additional verbs mention an apparently random assortment which usually includes some of the following: *add, admit, advise, announce, answer, assure, ask, claim, comment, complain, confess, convince, declare, demand, explain, indicate, inform, insist, invite, knew, mention, notify, order, persuade, point out, promise, propose, recommend, remark, remind, reply, replied that, report, require, shout, state, suggest, swear, teach, think, want to know, warn, whisper.*

***(2.b) Which tenses are used in textbook reported speech?*** The seven textbooks, first and foremost, generally instruct students to “backshift” both the main reporting verb and the embedded clausal (reported speech) verb. These instructions revolve around a system of “tense sequence rules” which tell students which exact past tense form to convert the reported speech verb into. The shifting of the tenses into a further past tense form is supposed to reflect the change in time between the initial utterance’s time frame and the current time frame and relationship to the reported speech.

Most of the textbooks acknowledge that sometimes other verb tense combinations are used to report speech. They generally present these other tense alternatives as exceptions to the

usual rule of tense backshifting. The textbooks describe certain grammatical and situational conditions that cause verb tenses not to conform to the typical past tense backshifting patterns, such as “backshifting often not applied when something is still present or true”; “backshifting not applied when the main (reporting) verb is in the present tense.” Yet by structuring their presentations in such a way, the textbooks often present a very complicated list of secondary rules for students to check.

Most of the textbooks hedged their secondary instructions a fair amount by saying that “usually” or “often” these alternative rules and exceptions apply. One alternative rule, however, is presented quite definitively by many textbooks. Three textbooks state outright that the reported speech embedded clause verb should not be backshifted to past tense when the main (reporting) verb is in the present tense, as in *Bob says he is sick*. So according to three textbooks, a Present-Past tense combination will never occur.

***(2.c) Do these textbook presentations of reported speech reflect authentic usage patterns?*** To answer this question, the results of the textbook analysis are compared to the corpus analysis. We discuss the extent to which the real language in the corpus followed the rules and guidelines in the grammar books. By objectively reporting and analyzing the results from both sources, a fair comparison can be made.

#### *Reporting Verb Authenticity*

In regards to which reporting verbs are used in reported speech, textbooks presentations reflect naturally-occurring usage patterns by endorsing *say* and *tell* as the most common reporting verbs. In regards to the “other” smaller reporting verbs, textbooks reflect naturally-occurring usage patterns by designating these verbs as secondary to *say* and *tell*, as, indeed, the corpus data confirms. However, in some ways, textbooks also do not reflect naturally-occurring usage patterns with these smaller verbs in that not all of the verbs listed in textbooks appeared in the corpus as reported speech.

Perhaps this discrepancy between some textbook information and authentic corpus findings involves the matter of prioritizing verbs according to register. For even though the corpus did find reported speech instances that used many different “other” verbs, not all of

these verbs were equal. As mentioned in the previous discussion, only a few of the “other” verbs could be recognized as prevalent enough to really matter (matter as in being numerous enough to study and find patterns within). Additionally, when a smaller, “other” verb was found to be frequent, its status was highly dependent on which register it was in (*agree*, *claim*, *indicate*, and *warn* in News; *ask* in Conversation). Therefore, textbook presentations of smaller “other” verbs could become more accurate by prioritizing the frequency of those verbs *and* including register variation.

### *Verb Tense Combination Authenticity*

With respect to verb tense combinations, textbook presentations reflect naturally-occurring language usage in some ways, and do not reflect naturally-occurring language usage in other ways. By instilling in students that the nature of reported speech has a Past-Past backshifting predisposition, textbooks teach students what the corpus reveals is the most common verb tense pattern in reported speech. So, in this way, textbooks accurately reflect naturally-occurring language by emphasizing the most common actual reported speech verb tense combination. However, textbooks also misrepresent naturally-occurring language usage when they claim that the past tense backshifting process does not occur when the main (reporting) verb is in the present tense. The corpus data reveals that, although the least common among the four potential verb tense combination categories, such a Present-Past sequence does indeed occur in both registers with both *say* and *tell*. These authentic corpus samples are examples of such occurrences:

... *but legal experts say that in trying to paper over deep philosophical divisions on the issue of racial preferences, the lawmakers further clouded many issues* (news)

... (a) *black film actor says police held a gun to his head* (news)

... *it says that he um, he did a few operas and he gave up on operas as soon as he heard Beethoven's operas ...* (conv)

By reinforcing the backshifting Past-Past tense combination as the norm, textbooks seem to imply that Past-Present and Present-Present forms are not very common, when, in fact, the corpus results find them not uncommon. Past-Present combinations range from 20 to 32 percent of reported speech in files, and Present-Present accounts for up to 14 percent of

reported speech in files. This sort of textbook presentation may not necessarily qualify as dangerously unauthentic, but it might be considered slightly misleading to students—especially advanced students—or just “uninformed.”

### *Authenticity of Reporting Verb Tense and Register*

Similarly, as mentioned earlier, one textbook (Fuchs & Bonner, 1995) asserted that a person is more likely to hear a present tense reporting verb in News. The corpus results follow this direction when *tell* is the reporting verb, but do not exhibit this characteristic when *say* is the reporting verb. Within the *say* occurrences, a Present-Present combination actually occurs more in Conversation than in News. So while we may associate a present tense reporting verb with News settings, it seems that we use this construction just as commonly in Conversation, but do not notice it as much.

### *Register Variation Authenticity*

One last issue, that of register variation, plays a crucial role in an assessment of textbook authenticity. Textbooks do not seem to address the impact of register variation in any way, and this represents a major oversight in describing naturally-occurring reported speech. The empirical corpus findings repeatedly demonstrated how both verb selection and tense combination often differed between the News and Conversation registers. So textbooks which present reported speech as conforming to the same rules in all contexts do not accurately depict authentic language usage.

### *Summary*

To sum up the question of textbook authenticity, some portions of the textbook presentations do accurately reflect naturally-occurring usage patterns, but other portions do not. Therefore, we must judge reported speech presentation in textbooks as not entirely authentic:

- Textbooks are accurate in presenting *say* and *tell* as the most common reporting verbs, but need to modify (specifically, prioritize) their presentations of the “other” smaller reporting verbs to reflect naturally-occurring language usage in line with empirical corpus data.
- Textbooks are authentic in that they present Past-Past backshifting as the most common verb



tense sequence, but are not authentic in their treatment of the other three tense combinations. Textbook characterizations of Past-Present and Present-Present are often misleading, and the textbook instructions regarding Present-Past are verifiably incorrect.

- Since textbooks do not address the issue of patterned register variation, they do not portray reported speech authentically in this area.

## Summary

This chapter outlined the quantitative findings from both the textbook analysis results and the corpus analysis results. Empirical corpus results have given new insights into the complicated behavior of reported speech and have shown how the behavior varies across registers. The basic findings from this study indicated that textbooks could represent reported speech more authentically—in terms of “other” reporting verbs used, verb tense combinations, and register variation.

Keep in mind that this discussion does not intend to assert so much that textbooks are currently wrong, but rather, such discussion of this study hopes to offer new insights for creating more informed textbook presentations.

Chapter five will further discuss major points from the study. We will expand the discussion to include implications and practical applications of the study’s results.

## CHAPTER 5. CONCLUSIONS

In these final pages I will try to succinctly reiterate and interpret how our current assumptions about reported speech (represented in grammar books) compare with the new information we have gained about reported speech (from the corpus study). This addresses the authenticity of textbook data compared to corpus results and how these results could influence grammar teaching. As both perspectives provide insight, and since both have potential to enrich our understanding of this grammatical topic, this summation aims to be a cooperative effort that can acknowledge the contributions of both textbooks and corpora.

This chapter will discuss key points from the study, including the research question implications, the study's limitations, and related future research.

### Key Points from this Study

Both the brief textbook study and the extensive corpus study resulted in substantial quantitative data about reported speech forms. Data from each analysis, however, represent vastly different approaches to studying grammatical topics. The textbook study becomes a summation of prescriptive grammar rules, while the corpus study becomes a descriptive grammar by its very nature. (Or, at best, the textbook study represents the current state of textbook authors' intuitions—or sincere, well-intended, best guesses and observations—about reported speech behavior.) Since prescriptive grammar manifests from grammarians' assumptions about language, and since descriptive grammar manifests from real data gathered from naturally-occurring language, the premise of the textbook-corpus comparison sets up a challenge to investigate how authentically textbooks instruct students in reported speech. Remember, however, that the issue is not textbook accuracy, but rather textbook authenticity. Of course, no textbook author intends to give inaccurate information to English language

learners. But without empirical data, authors can only rely on intuitions. The following data comparisons should reveal the relationship between prescriptions about proper reported speech usage (from textbooks) and the reality of actual reported speech usage (from corpora).

### **Incorporating Both Textbook Information and Corpus Information**

My assessment of ESL/EFL grammar textbooks concludes that most textbooks present confusing, sometimes conflicting or incomplete, information about reported speech to students. I believe this happens because textbook authors themselves are somewhat unsure and confused about reported speech. This confusion exists because naturally-occurring reported speech behavior, as shown through this empirical corpus study, actually is complicated, confusing, and sometimes conflicting! People say and do strange things with this language function! Therefore, it seems that we could certainly use a more informed description of this language phenomenon.

It seems that a more informed description may come from merging grammarian's assumptions and prescriptions with real-life descriptions from corpora's naturally-occurring, concrete evidence. If we suppose that the current grammar textbook descriptions are evidence of what we currently believe about reported speech, then we should be willing to accept that an empirical corpus study can further our understanding. Corpus linguistics becomes a tool that allows us to isolate and examine this interesting language function—reported speech—in action. We can observe how it occurs across many different speakers, situations, places, times, etc. in a highly objective way. This enables us to check our assumptions and tweak them if necessary. We can then strive to unify these findings and achieve a more accurate, informed description of reported speech.

Below are some key points that surfaced from that type of comparison in this particular study:

### **Implications of Research Questions and Findings**

This section reassesses textbooks in light of the corpus findings and discussion, and gives suggestions for textbook modifications and improvements. This section also discusses SLA-principled grammar instruction in terms of reported speech authenticity and corpus-based

materials.

### **Textbooks vs. corpora**

As stated initially, this study does not aim to criticize textbooks. Instead, this study is based on the premise that reported speech is a particularly complicated and troublesome area of grammar for students. Grammar textbooks attest to this complexity when they each, often, give different conflicting descriptions and instructions about reported speech to students. So this textbook-corpora comparison does conjure up attitudes regarding prescriptive vs. descriptive grammar and language use in general. But we will first approach the issue from a standpoint of enabling textbooks to use corpus findings to incorporate reported speech descriptions which are more helpful to students.

### ***Implications for textbook presentation of verb tense combinations***

Perhaps most noteworthy in the study is the extent to which non-traditional reported speech tenses pervade the corpus results and those implications. If the traditional backshifting of tenses accounts for no more than 67%, at the absolute most, of reported speech occurrences (and a mere 51% at its least), then why do textbooks focus on this Past-Past combination as the one, pure form of reported speech? Shouldn't instructions to students include the other tense combination alternatives, particularly the Past-Present forms which occur in almost one-third of the occurrences in some files? Although the textbooks discussed Past-Present combinations in terms of "exceptions," these corpus results call for Past-Present examples alongside the prominent Past-Past examples.

The textbooks currently do a good job of thoroughly describing the traditional Past-Past tense combination. They also cover the Past-Present combination a fair amount through the "exceptions to the norm" that they present. However, the corpus data results seem to indicate that the non-traditional verb tenses, particularly Past-Present, deserve more status than just "exceptions." Patterns do surface within all of the verb tense combinations. These other verb tenses are not random exceptions. So it seems that an accurate depiction of reported speech in ESL grammar books should include discussions *and examples* of all potential verb tense combinations that can and do occur in reported speech—particularly with further advanced ESL

students.

Of course, textbooks should prioritize the amount of emphasis given to each combination, and with corpus data, textbooks have the ability to make judgments regarding appropriate focus amounts. So Past-Past combinations could still be the primary focus, but probably 20 to 30 percent of the focus and samples should involve Past-Present combinations, and 10 to 15 percent should involve Present-Present and Present-Past combinations. These “alternative” tense combinations should probably be given independent status as well, indicating that they are legitimate verb forms and not mere exceptions to Past-Past—not deviations or flaws of the standard pure form.

On the other hand, elaborate grammatical explanations are not recommended for lower level ESL students. Focus-on-form research even acknowledges that extensive attention to grammatical detail is not appropriate with low proficiency students (VanPatten, 1988). Therefore, textbooks should present empirical data appropriately to various levels of students. Naturally-occurring reported speech patterns can be presented without overwhelming beginning-level students. More explanation can be given for higher-level students.

***Implications for textbook presentation of reporting verbs*** Given the more informed insights into reporting verbs used, textbooks should revise their presentations to concentrate more on the frequent reporting verbs. Perhaps a greater distinction should be made between *say* and *tell* as well. These two main reporting verbs should not be presented as equal when, in reality, *say* is much more common. Samples and example practice exercises should reflect these informed verb priorities too.

***Implications for textbook presentation of register variation*** Register variation surfaces as an issue in many places in reported speech. At a general level, the corpus data shows that reported speech will occur much more often in News than in Conversation—at least three times more often in News than Conversation. More specific register variation influences particular reported speech behaviors. For instance, the reporting verb choice depends, to a large extent, on register. News uses a larger variety of reporting verbs and also is more likely to use such non-say-tell verbs than Conversation is. And register variation

significantly affects tense combinations. The corpus results indicate that both *say* and *tell* exhibit more “unusual,” atypical tense combinations in Conversation than in News, meaning that News uses proportionally more traditional Past-Past tense combinations than does Conversation. And register variation seems as if it might be one of the key factors that influences unusual reported speech behavior such as Present-Past combinations.

Textbooks need to incorporate register variation. In light of the empirical findings, textbooks may even want to consider explaining these differences explicitly in their directions. It seems that students will encounter reported speech much more frequently in News than in Conversation, so textbook explanations, samples, and practice activities should reflect this. Why should students practice as many conversational reported speech instances as news reported speech instances when reported speech in News is so much more prevalent? Or perhaps this justifies listening to and practicing conversational reported speech (since that is the context in which most students will probably need to produce reported speech), but justifies exposure to even more news reported speech (since students will probably hear and need to interpret much more reported speech in this setting than they will have to produce).

Register variation influences the type of reported speech that students should learn as well. If News reported speech is more traditional and conventional (using more Past-Past constructions), then textbooks can use corpus findings to present more of the non-Past-Past formats in conversational contexts—not News.

### **SLA-principled grammar teaching in light of findings**

How can these findings refine grammar teaching materials in accord with SLA principles? For one thing, empirical corpus results will enable textbooks to encourage students to learn more independently, more authentically, and with more analysis which should foster further negotiation of meaning.

Teachers can tailor empirical corpus findings and data to fit student needs and address student questions. Teachers can manipulate the corpus data to provide optimal “comprehensible input” in a communicative setting. Teachers can also implement focus on form and consciousness-raising by highlighting and guiding students through the relevant corpus samples. Corpus samples can be selected to illustrate particular grammatical features

(such as tense combination) or factors (such as situational context) which influence the reported speech verb form choice. Students can analyze WHY speakers might have made certain verb and tense choices. Then students and teachers can have confidence in their speculations since the samples they look at come from authentic corpus samples and repeatedly illustrate patterns. These sorts of student analysis activities would benefit students by combining empirical corpus findings with SLA research findings.

From this study's results, we can also assume that students will encounter reported speech more often in the context of News (since the corpus results showed that reported speech is ~3.5 times more common in News than Conversation). If reported speech instances occur ~8 out of 10 times in News, and ~2 out of 10 times in Conversation, it seems logical, therefore, that student exposure to reported speech should involve a lot of News. Certainly, if focusing on these two registers with students, at least half of the samples and exercises in textbooks should come from News contexts. Given this empirical information, a teacher would also want to create as natural of a language environment as possible by focusing presentations and practice of reported speech with students around News (in a general grammar class—not a conversation class, of course). While a teacher should certainly address and have students practice conversational reported speech too, the largest quantity of emphasis should probably be on News reported speech to reflect real life language usage. Additionally, further corpus studies might reveal other registers besides News with high reported speech frequencies.

### **Reported speech grammar teaching and materials in general**

ESL students and teachers stand to benefit the most through more informed descriptions of authentic language use in their grammar textbooks and teaching materials:

*Reported speech can be made less confusing to students and inherently more accurate by prioritizing the grammar book presentations according to the frequent behaviors found in authentic usage* This can be done with the reporting verbs as well as the tense combinations. Textbooks need not overwhelm beginning students with lengthy lists of “other” reporting verbs when those verbs hardly ever occur in reality.

Textbooks need not have equal numbers of “say examples” and “tell examples” when *say* is much more common than *tell*. Textbooks do not need to classify all non-Past-Past tense combinations as “exceptions.” Proportionate samples of the various tense options can be presented, prominently, as legitimate forms in their own right.

*Textbooks should address register variation to a greater extent in their presentations of reported speech* If textbooks would integrate explanations about register variation, their depictions of reported speech would immediately become more accurate. It might also make textbook presentations less confusing by eliminating the need for so many exceptions to one universal rule. Instead, two types of reported speech behaviors could be distinguished: reported speech patterns in Conversation or other spoken registers and reported speech patterns in News or other written registers. This sort of description would give students a more accurate representation of the naturally occurring reported speech that they will encounter in real life. Of course, further corpus studies would be necessary in order to describe reported speech behavior in other registers.

### **Limitations**

Findings from this study are limited by some research factors. The small number of textbooks (seven) limits the ability to generalize findings to all ESL textbooks. Yet the textbooks used in the textbook analysis were chosen selectively; they represent the most thorough texts found to be available. Also, this study dealt with two registers only—Newspapers and Conversation. This again limits our ability to generalize findings to all of naturally occurring speech. Additionally, some of the “other” verbs were extremely rare in the corpus, and so these small numbers inhibit our ability to analyze naturally-occurring reported speech behavior with those verbs. Lastly, the fact that only one researcher carried out the data analysis increases the chance of human error. Even though the hand-editing judgement calls and verb tense categorizing were extremely straightforward (see the Methodology chapter), additional participants would enable us to have more confidence in the quantitative results’ validity.

But of course, these limitations could be accounted for with related future research



endeavors.

### **Future Research**

This study has also revealed areas which would benefit from further investigation:

At a general level, more registers could be incorporated into a similar study. Additional research to examine how naturally-occurring reported speech occurs in Spoken News, Fiction, and other registers would be quite interesting.

Since this study ultimately focused on *say* and *tell* as its major verbs, future research to study other reporting verbs in more depth would be beneficial. These less frequent “other verbs” might surface in other registers as well and exhibit different patterns than in News and Conversation. It would be useful to find another way to search for reporting verbs as well (besides relying on textbook suggestions and intuition). Perhaps this would give clues to why certain reporting verbs are used in certain situations.

Relatedly, there is a need for more in-depth analyses of verb choices and tense combination choices. This study’s corpus data cannot sufficiently show us enough of the situational context and relational variables to be able to say what external, or even grammatical, conditions cause specific language behavior. We cannot know, with much certainty anyway, what exactly made a speaker use “says” instead of “said” to report the speech. But a study could be designed to explore this in the future. Perhaps this would entail gathering data about hearers’ and readers’ perceptions of various reported speech forms. Maybe this could involve interviewing writers about the reported speech choices they made.

Also, one of the textbooks (Bland, 1996) identified certain verbs as having unique behavior involving the subjunctive tense. The instructions claim that the reporting verbs *ask*, *require*, *insist*, *demand*, *suggest*, *recommend*, *advise*, and *propose* all will have their subsequent clausal verbs occur in the simple form (“I recommend that he stay”; “They suggested that she take a vacation”) (p. 468). These assumptions could be checked with empirical corpus evidence.

Finally, more textbooks could be analyzed and, eventually, more textbooks could be written. Investigating the perceptions that ESL/EFL students hold about reported speech might supply useful data for textbook writing too.

## Summary

Results from this corpus study contribute to the process of furthering our knowledge and insight into naturally-occurring reported speech. This sort of corpus study demonstrates the sort of unique perspective that empirical language descriptions result in. Although the information gained from such research may not be simple or clear-cut, it undoubtedly gives an authentic description of our linguistic reality.

One extremely clear finding from the corpus study is that reported speech varies significantly between registers. Both verb choice and tense choice happen differently in News than in Conversation. This issue, alone, might represent the most substantial finding in the study for two basic reasons: 1) the most definitive results from this study show evidence of patterned register variation 2) most textbooks do not even mention register variation as an issue. Since the results point to register as a factor that most certainly influences reported speech behavior, textbooks owe it to students to address such issues. This finding, and similar results from this study, can be useful in presenting models to students and in predicting the likelihood of specific reported speech occurrences.

Hence, information gained from this study could be applied for numerous purposes. Most importantly, they can all relate to how these empirical findings about reported speech can refine grammar teaching materials in accordance with SLA principles—either from this present study's insights or with insights from future research in this area.

## APPENDIX A. SAMPLE KWIC FILES

The computer program displayed the corpus data in KWIC (Key Word In Context) format. The data used for the empirical analysis appeared in this format:

Item = asked  
File = 06051100.SJ  
that they will run for the long - term seat  
. ; to measure candidates ' relative strength , field  
-----> asked  
those polled if they had an opinion of potential  
candidates and whether that opinion was favorable or unfavorable .  
; Of the 977 people surveyed Feb . 7 -  
12 , 86 percent of all voters knew who Feinstein

Item = asked  
File = 06051189.SJ  
drained today , " he scribbled on a message slip  
in october 1986 , according to court documents . ellison  
-----> asked  
a co - worker to tell Gray to leave  
her alone , to no avail . " i know  
that you are worth knowing with or without sex ,  
" he wrote in a single - spaced , three

Item = asked  
File = 06051322.SJ  
stories were published in the feb . 13 extra .  
ten minutes before the ceremony was to start , i  
-----> asked  
my best man to peek into the church and  
" count the house . " He rattled the knob  
of the connecting door and then told me that it  
was locked . I thought he was kidding , until

Item = asked  
File = 06077012.SJ  
, " says spokeswoman dianne mathews . early risers are  
served fresh eggs . ; ( check ) american reporters  
-----> asked  
Lech Walesa if being president of Poland was everything  
he imagined . His answer : " oh no ,  
it 's worse . I knew it would be difficult  
and ungratifying , but it is worse and even more

## APPENDIX B. DETAILED VERB TENSE COMBINATIONS

### Past-Past:

past --> past  
 past --> past progressive  
 past --> past perfect  
 past --> past perfect passive  
 past --> past passive  
 past --> M-past + present  
 past --> DID + present  
 past --> M-past + present progressive  
 past --> M-past + present perfect  
 past --> M-past + present passive  
 past perfect --> past  
 past perfect --> M-past + present  
 past passive --> past  
 DID + present --> past  
 M-past + present --> past

### Past-Present:

past --> present  
 past --> present progressive  
 past --> present perfect  
 past --> present perfect progressive  
 past --> present passive  
 past --> DO/DOES + present  
 past --> M-present + present  
 past --> M-present + present perfect  
 past --> M-present + present passive  
 past perfect --> present  
 M-past + present perfect --> present  
 DID + present --> present

### Present-Past:

present --> past  
 present --> M-past + present  
 present --> M-past + present perfect  
 present perfect --> past  
 present perfect --> M-past + present  
 DO/DOES + present --> past

### Present-Present:

present --> present  
 present --> DO/DOES + present  
 present --> present progressive  
 present --> present perfect  
 present --> present perfect passive  
 present --> present passive  
 present --> M-present + present passive  
 present perfect --> present  
 present perfect --> present perfect  
 present perfect --> M-present + present  
 present perfect --> M-present + present passive

\* M-past could be *could, might, would, should*

\* M-present could be *can, may, will, shall*

## APPENDIX C. RESULTS WITH “OTHER VERBS”

### Say

#### SAY – American News

Word Count: 3,267,337 (size of entire Am.News corpus)

Size of Results: 3,593 pages of potential reported speech; 17,961 occurrences

Proportion of R.S. in results: = **76.5%** of “say” instances in Am.News results were r.s.

Frequency: **4205** r.s. occurrences **per million words** (per million normed counts)

#### SAY – American Conversation

Word Count: 2,071,994 (size of entire Am.Conv corpus)

Size of Results: 845 pages of potential reported speech; 4,222 occurrences

Proportion of R.S. in results: **49.6%** of “say” instances in Am.Conv results were r.s.

Frequency: **1011** r.s. occurrences **per million words** (per million normed counts)

### Tell

#### TELL – American News

Word Count: 3,267,377 (size of entire Am.News corpus)

Size of Results: 314 pages of potential reported speech; 1,570 occurrences

Proportion of R.S. in results: **70.8%** of “tell” instances in Am.News results were r.s.

Frequency: **340** r.s. occurrences **per million words** (per million normed counts)

#### TELL – American Conversation

Word Count: 2,071,994 (size of entire Am.Conv corpus)

Size of Results: 296 pages of potential reported speech; 1,479 occurrences

Proportion of R.S. in output file: **44.4%** of “tell” instances in Am.Conv results were r.s.

Frequency: **317** r.s. occurrences **per million words** (per million normed counts)

### Others

#### OTHER SAY-type VERBS –American Conversation

Verb	Positive / Total Instances/ Instances	Proportion %	Occurrences Per Million
Add	0/95	0%	
Admit	2/10	20%	
Agree	4/45	8%	
Announce	2/10	20%	
Answer	0/55	0%	
Claim	8/15	<b>53%</b>	.....~ 4/mil.
Complain	2/5	40%	
Comment	2/5	40%	
Confess	0/1	0%	
Declare	2/5	40%	
Explain	2/50	4%	
Indicate	2/10	40%	
Mention	17/70	24%	..... ~8/mil.
Remark	0/0	0%	
Report	1/10	10%	
Reply	0/1	0%	
Show?	0/0	0%	
Swear	17/30 ??	<b>57%</b>	.....~8/mil.
Whisper	0/0	0%	

... continued on next page...

OTHER SAY-type VERBS – American News

Verb	Positive / Total Instances/ Instances	Proportion %	Occurrences Per Million
Add	108/354	31%	33/mill.
Admit	76/192	40%	23/mill.
<b>Agree</b>	<b>129/192</b>	<b>67%</b>	<b>39/mill.</b>
Announce	255/629	41%	78/mill.
Answer	0/48	0%	
<b>Claim</b>	<b>167/306</b>	<b>55%</b>	<b>51/mill.</b>
Complain	104/306	34%	32/mill.
Comment	4/30	13%	
Declare	44/216	21%	13/mill.
Explain	31/114	27%	9/mill.
<b>Indicate</b>	<b>197/300</b>	<b>66%</b>	<b>60/mill.</b>
Mention	4/126	3%	
Remark	2/12	16%	
Report	51/283	18%	78/mill.* every 5th inst. from 1
Reply	16/24	67%	5/mill.
Show?	0/47	0%	
Swear	8/14	57%	2/mill.
Whisper	41/121	34%	

OTHER TELL-type VERBS – American Conversation

Verb	Positive / Total Instances/ Instances	Proportion %	Occurrences Per Million
Advise	1/1	100%	0.5/mill.
Ask	25/94	27%	61/mill. * every 5th inst. from 1
Assure	3/4	75%	1/mill.
Convince	3/8	38%	
Inform	1/3	33%	
Notify	0/1	0%	
Persuade	0/0	0%	
Promise	3/13	23%	
Remind	1/55	2%	
Teach	6/81	7%	
Warn	2/3	67%	1/mill.

OTHER TELL-type VERBS – American News

Verb	Positive / Total Instances/ Instances	Proportion %	Occurrences Per Million
Advise	18/49	37%	
<b>Ask</b>	<b>50/104</b>	<b>48%</b>	<b>76/mill. * every 5th inst. from 1</b>
Assure	29/47	62%	9/mill.
Convince	12/24	50%	4/mill.
Inform	15/28	54%	5/mill.
Notify	15/26	58%	5/mill.
Persuade	11/14	79%	3/mill.
Promise	11/34	32%	8/mill. * 2 inst. per page from 1
Remind	9/33	27%	
Teach	0/36	0%	* 2 instances per page from 1
<b>Warn</b>	<b>45/56</b>	<b>80%</b>	<b>34/mill. * 2 inst. per page from 1</b>

## REFERENCES

- Aijmer K. & Altenberg, B. (Eds.). (1991). *English corpus linguistics: Studies in honour of Jan Svartik*. Singapore: Longman.
- Biber, D. (2001, February). *Academic discourse: Corpus-based perspectives*. Plenary session presented at the American Association of Applied Linguistics Annual Conference, St. Louis, MO.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Biber, D. & Finegan, E. (1991). On the exploitation of computerized corpora in variation studies. In Aijmer K. & Altenberg, B. (Eds.). (1991). *English Corpus Linguistics: Studies in honour of Jan Svartik* (pp. 204-220). Singapore: Longman.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow, Essex, UK: Pearson Education.
- Bland, S. K (1996). *Intermediate grammar: From form to meaning and use* (pp. 461-475). Oxford: Oxford.
- Blyth, C. Jr., Recktenwald, S., & Wang, J. (1990). I'm like, "say what?!": A new quotative in American oral narrative. *American Speech* 65 (3), 215-227.
- Buttny, R. & Williams, P. L. (2000). Demanding respect: The uses of reported speech in discursive constructions of interracial contact. *Discourse and Society*, 11 (1), 109-133.
- Comrie, B. (1986). *Tense in indirect speech*. *Folia Linguistica*, 20, 265-296.
- Conrad, S.M. (1999). The importance of corpus-based research for language teachers. *System* 27, 1-18.
- Conrad, S.M. (2000). Will corpus linguistics revolutionize grammar teaching in the 21st century? *TESOL Quarterly*, 34 (3), 548-560.
- Cook, V. (1999). Using SLA research in language teaching. *International Journal of Applied Linguistics*, 9 (2), 267-284.
- Declerck, R. & Tanaka, K. (1996). Constraints on tense choice in reported speech. *Studia Linguistica*, 50 (3), 283-301.
- Doughty, C. & Varela, E. (1998). Communicative focus on form. In C. Doughty & J. Williams. (Eds.). (1998). *Focus on form in classroom second language acquisition* (pp. 114-138). New York: Cambridge University Press.
- Eastwood, J. (1999). *Oxford practice grammar with answers* (pp. 318-329). Hong Kong: Oxford University Press.

- Elbaum, S. N. (2001). *Grammar in context* (3rd ed.) (pp. 362-385). Boston: Heinle & Heinle.
- Ellis, R. (1994). *The study of second language acquisition*. Hong Kong: Oxford University Press.
- Ferrara, K. & Bell, B. (1995). Sociolinguistic variation and discourse function of constructed dialogue introducers: The case of *be + like*. *American Speech* 70 (3), 265-290.
- Fotos, S. & Ellis, R. (1991). Communicating about grammar: A task-based approach. *TESOL Quarterly*, 25 (4), 605-628.
- Fox, G. (1998). Using corpus data in the classroom. In B. Tomlinson (Ed.). *Materials development in language teaching* (pp. 25-43). Cambridge: Cambridge University Press.
- Fuchs, M. & Bonner, M. (1995). *Focus on grammar: A high-intermediate course for reference and practice* (pp. 271-328). Addison-Wesley; Longman.
- Goodell, E.W. (1987). Integrating theory with practice: An alternative approach to reported speech in English. *TESOL Quarterly*, 21 (2), 305-325.
- Huddleston, R. (1989). The treatment of tense in indirect speech. *Folia Linguistica*, 23 (3), 335-340.
- Jacobs, G. (1999). Tense choice in English past reported speech. *Leuvense Bijdragen*, 88, 51-67.
- Johns, T. (1994). From printout to handout: Grammar and vocabulary teaching in the context of Data-driven Learning. In T. Odlin (Ed.). *Perspectives on pedagogical grammar* (pp. 293-313). New York: Cambridge University Press.
- Johnstone, B. (1987). 'He says ... so I said': verb tense alternation and narrative depictions of authority in American English. *Linguistics* 25, 33-52.
- Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. London: Longman.
- Lightbown, P.M. (1987). Classroom language as input to second language acquisition. In Pfaff, C. W. (Ed.). *First and second language acquisition processes* (pp. 169-187). Cambridge: Newbury House Publishers.
- Lightbown, P. M. & Spada, N. (1993). *How languages are learned*. Hong Kong: Oxford University Press.
- Lightbown, P.M. & Spada, N. (1990). Focus-on-form and corrective feedback in communicative language teaching: Effects on second language learning. *Studies in Second Language Acquisition*, 12, 429-448.
- Little, D. (1994). Words and their properties: Arguments for a lexical approach to pedagogical grammar. In T. Odlin (Ed.). *Perspectives on pedagogical grammar* (pp. 99-122). New York: Cambridge University Press.



- Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. De Bot, R. Ginsberg, & C. Kramsch (Eds.). *Foreign language research in cross-cultural perspective* (pp. 39-52). Philadelphia: John Benjamins Publishing.
- Long, M. H. (1985). Input and second language acquisition theory. In S. M. Gass, & C. G. Madden. (Eds.). (1985). *Input in second language acquisition* (pp. 377-393). Rowley, Mass.: Newbury House.
- Long, M. H. & Robinson, P. (1998). Focus on form: Theory, research, and practice. In C. Doughty & J. Williams. (Eds.). (1998). *Focus on form in classroom second language acquisition* (pp. 15-41). New York: Cambridge University Press.
- Murphy, R., Altman, R. & Rutherford, W. E. (1989). *Grammar in use: Reference and practice for intermediate students of English* (pp. 90-93). New York: Cambridge University Press.
- Partington, A. (1998). *Patterns and meanings: using corpora for English language research and teaching*. Philadelphia: John Benjamins Publishing.
- Pica, T. (1994). Questions from the language classroom: Research perspectives. *TESOL Quarterly*, 28 (1), 49-79.
- Pica, T. & Doughty, C. (1985). The role of group work in classroom second language acquisition. *Studies in Second Language Acquisition* 7, 233-248.
- Raimes, A. (1998). *Grammar troublespots: An editing guide for students* (2nd ed.) (pp. 141-145). New York: Cambridge University Press.
- Rings, L. (1992). Authentic spoken texts as examples of language variation: Grammatical, situational, and cultural teaching models. *IRAL*, 30 (1), 21-33.
- Romaine, S. & Lange, D. (1991). The use of *like* as a marker of reported speech and thought: A case of grammaticalization in progress. *American Speech* 66 (3), 227-279.
- Rutherford, W. E. (1987). The meaning of grammatical consciousness-raising. *World Englishes*, 6 (3), 209-216.
- Rutherford, W. E. (1987). *Second language grammar: Learning and teaching*. Singapore: Longman.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Hong Kong: Oxford University Press.
- Spada, N. M. (1987). Relationships between instructional differences and learning outcomes: A process-product study of communicative language teaching. *Applied Linguistics*, 8, 137-161.
- Stubbs, M. (1996). *Text and corpus analysis: Computer assisted studies of language and culture*. Oxford: Blackwell.
- Swain, M. (1998). Focus on form through conscious reflection. In C. Doughty & J. Williams. (Eds.). (1998). *Focus on form in classroom second language acquisition*

(pp. 64-81). New York: Cambridge University Press.

- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass, & C. G. Madden. (Eds.). (1985). *Input in second language acquisition* (pp. 235-253). Rowley, Mass.: Newbury House.
- Tagliamonte, S. & Hudson, R. (1998). Be like et al. beyond America: The quotative system in Canadian and UK youth. *Cahiers linguistiques d'Ottawa*, 26, 55-69.
- Thewlis, S. H. (2001). *Grammar dimensions* (3rd ed.) (pp. 400-416). Boston: Heinle & Heinle.
- Thompson, G. (1996). Voices in the text: Discourse perspectives on language reports. *Applied Linguistics*, 17 (4), 501-530.
- VanPatten, B. (1988). How juries get hung: Problems with the evidence for a focus on form in teaching. *Language Learning*, 38 (2), 243-260.
- Yip, V. (1994). Grammatical consciousness-raising and learnability. In T. Odlin (Ed.). *Perspectives on pedagogical grammar* (pp.123-138). New York: Cambridge University Press.
- Yule, G. & Mathis, T. (1992). The role of staging and constructed dialogue in establishing speaker's topic. *Linguistics* 30, 199-215.
- Yule, G., Mathis, T., & HopKins, M. F. (1992). On reporting what was said. *English Language Teaching Journal*, 46 (3), 245-251.

## ACKNOWLEDGEMENTS

I owe thanks to a number of people who helped with and contributed to this study.

- Dr. Susan Conrad, my major professor, who worked so diligently and patiently with me throughout this project—even in Italy and Portland! Thank you for introducing me to corpus linguistics, writing the computer programs, giving me so many good ideas (and words!) for this write-up, and thinking that grammar is interesting to analyze like I do. This thesis would not have been possible without your guidance, and I have learned a great deal from working with you!
- Dr. John Hagge and Dr. Kathy Leonard, committee members—your comments at the final oral exam meeting and on written drafts were very helpful
- Federica Barbieri, for being a valuable resource to discuss reported speech with and for help with some article citations; De Zhang, and my other Iowa State TESL/AL classmates
- Iowa State's TESL/AL professors, ESL staff, & the Intensive English and Orientation Program instructors and resources
- Marcus Eckhardt, patient and supportive throughout the whole process ... :)