

## INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

**The quality of this reproduction is dependent upon the quality of the copy submitted.** Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

# UMI

A Bell & Howell Information Company  
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA  
313/761-4700 800/521-0600



Not Another White Elephant!  
Children's Understanding Of Unfamiliar  
Phrasal Idioms.

Or,  
Paralysis By Analysis?

by

Glenn A. Johnson

A dissertation submitted in partial fulfillment of  
the requirements for the degree of

Doctor of Philosophy

University of Washington

1998

Approved by Lesley B. O'Way  
Chairperson of Supervisory Committee

Virginia W. Berninger

Ph. S. D. L.

Carol S.

Program Authorized  
to Offer Degree Speech and Hearing Sciences

Date 12/15/98

**UMI Number: 9916676**

**Copyright 1998 by  
Johnson, Glenn A.**

**All rights reserved.**

---

**UMI Microform 9916676  
Copyright 1999, by UMI Company. All rights reserved.**

**This microform edition is protected against unauthorized  
copying under Title 17, United States Code.**

---

**UMI**  
**300 North Zeeb Road**  
**Ann Arbor, MI 48103**

Copyright 1998

Glenn A. Johnson

## Doctoral Dissertation

In presenting this dissertation in partial fulfillment of the requirements for the Doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to University Microfilms, 1490 Eisenhower Place, P.O. Box 975, Ann Arbor, MI 48106, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."

Signature William A. Johnson  
Date 12/15/98

University of Washington

Abstract

Not Another White Elephant!  
Children's Understanding of Unfamiliar  
Phrasal Idioms.

Or,  
Paralysis by Analysis?

by Glenn A. Johnson

Chairperson of the Supervisory Committee:  
Professor Lesley B. Olswang  
Department of Speech & Hearing Sciences

This research had three specific aims related to children's comprehension of idioms: (1) to reexamine the impact of analyzability on children's comprehension of unfamiliar idioms; (2) to explore the patterns of developmental progression associated with verb phrase, noun phrase, adjective phrase, and syntactically anomalous idioms; and, (3) to identify any meaningful interactions between analyzability and form.

Thirty-six typically developing children, 12 each in grades two, five and eight, participated in the experimental tasks. Idioms were presented to students in short stories constructed to support a figurative interpretation of the idiom. Comprehension was assessed through the use of two response tasks: (1) an item-selection task, and (2) an item-explanation task (i.e., having students explain what they thought each idiom meant in their own words).

The outcomes indicated that analyzability does impact children's comprehension of idioms. However, analyzability interacts with response type and with idiom form. Students were more successful in explaining the figurative meanings of analyzable idioms in general, and were more successful in explaining verb phrase idioms than noun or adjective phrase idioms. Conversely, students were generally more successful in selecting the figurative

paraphrase for unanalyzable than analyzable idioms, though this effect was not significant for verb phrase idioms. The results indicate that idiom comprehension processes in children are best considered from within a discourse processes framework.

## TABLE OF CONTENTS

	Page
List of Figures .....	ii
List of Tables .....	iii
Chapter 1: Introduction .....	1
Objectives .....	2
Historical Overview .....	4
Review of the Literature .....	8
Framework .....	9
Theories of Idiom Processing and Lexical Representation .....	14
Studies of Typical Development .....	27
Summary .....	45
Statement of the Problem .....	48
Chapter 2: Methods .....	50
Chapter 3: Results .....	69
Chapter 4: Discussion .....	95
Conclusions .....	119
References .....	121
Appendix A: Instructions for Idiom Analyzability Ratings .....	130
Appendix B: Instructions for Idiom Familiarity Ratings .....	132
Appendix C: Stimuli .....	133
Appendix D: Instructions to the Participants .....	141

## LIST OF FIGURES

Number		Page
1.	Developmental Trends, Early Studies .....	28
2.	Effect of Response Tasks .....	32
3.	Impact of Context on Idiomatic Comprehension .....	32
4.	Forced-choice Responses for VP, NP and ADJP idioms .....	83
5.	Idiomatic Explanations for VP, NP and ADJP Idioms .....	83
6.	Forced-choice, SA-SU x Idiom Form x G2 - G8.....	88

## LIST OF TABLES

Number		Page
1.	Theories of Idiom Processing, Lexical Representation and Idiomatic Flexibility.....	14
2.	Effects of Analyzability on Comprehension in Typically Developing Children Reported in 8 Studies.....	38
3.	Effects of Familiarity on Comprehension in Typically Developing Children and Adults Reported in 8 Studies..	43
4.	Participant Demographic Data, by Grade.....	51
5.	Students Clinical Evaluation of Language Fundamentals-R and Wechsler Intelligence Scale for Children-R performance data by grade.....	53
6.	Idioms used in the Comprehension Tasks.....	57
7.	Response Categories and Criteria for the Explanation and Forced - Choice Comprehension Tasks.....	58
8.	Procedural Reliability: Judges Averaged Ratings for Prosody and Use of Verbal Reinforcement .....	63
9.	Procedural Reliability: Judges Averaged Ratings for Prosody and Use of Verbal Reinforcement .....	65

10.	Mean Number and Mean Percentage of Correct Item-Selection Responses for SA and SU Idioms by Grade with Post Hoc Significant Differences for Grade.....	71
11.	Analysis of Variance for Students' Item-Selections of SA and SU Idioms (ANOVA1).....	72
12.	Mean Number and Mean Percentage of Correct Explanations of SA and SU Idioms by Grade with Scheffe' Post Hoc Significant Differences for Grade.....	73
13.	Analysis of Variance for Students' Explanations of SA and SU Idioms (ANOVA2).....	74
14.	Analysis of Variance for Students' C-I + F-O Explanations of SA and SU Idioms (ANOVA3).....	76
15.	Mean Percentage of Correct - Idiomatic (C-I), Figurative - Other (F-O), and C-I + F-O Explanations of SA and SU Idioms by Grade.....	77
16.	Mean Percentage of Correct Item-Selection Responses for Each Idiom Form by Grade.....	80
17.	Mean Number and Mean Percentage of Correct Explanations for VP, NP and ADJP Idioms, by Grade.....	82

18.	Mean Percentage of Correct Choices for SA and SU Idioms, by Idiom Form and Grade.....	87
19.	Mean Percentage of Correct Explanations for SA and SU Idioms, by Idiom Form and Grade.....	89
20.	Mean Percentage of Correct Item-selection and Explanation Task Responses, G2 - G8 .....	91
21.	Multivariate Analysis of Variance for Interactions between Analyzability (SA - SU), Response Task (I/S - Exp), Idiom Form (Form) and Grade(MANOVA4).....	93
22.	Significant Main Effects for Analyzability, Form, Grade, and Their Interactions, by Response Condition .....	97

## ACKNOWLEDGEMENTS

I would like to thank the chair of my committee and tireless editor, Dr. Lesley Olswang, for her support, encouragement and patience throughout my U.W. career. I don't always make it easy. I would also like to thank Dr. Phillip Dale, Dr. Carol Stoel-Gammon, and Dr. Virginia Berninger for their comments, patience and wisdom in this project. I want to thank my entire committee for their good humor in bearing up under the nearly endless bombardment of idioms and other troupes that I tucked away in our conversations and writings.

Thanks to the students, staff and parents in the Shoreline School District, and at Our Lady of Guadalupe School in Seattle. It was a real pleasure to get to know the students who sat and listened, and thought and answered.

The assistance of Susan Coon-Donovan and Prudence Cromwell in recognizing and scoring the explanations provided by the creative students who taught us a lot about idiom meanings, is gratefully acknowledged.

Many thanks to Julee Neuhart, head of the Speech-Language Pathology program for the Seattle School District, for her support and flexibility in helping me find the time to finish this project this fall.

## DEDICATION

This work is dedicated to the double helix crew: To Cate, Jess and Prudence, who were encouraging and loving throughout this self-imposed in-house detention; and to my parents Al and Cass Johnson, who have supported each and every one of my endeavors.

Thanks beyond words.

## CHAPTER ONE

### Introduction

"*Keep my eyes peeled* for coyotes? Daddy, how can I peel my eye?"

"*A grain of salt*? How can I take a grain of salt - I don't even know what that is".

These two quotes - the first from a four year-old, the second from a six year-old, illustrate what many parents and developmental psycholinguists intuitively take *to be true* -- that the development of children's comprehension of idioms is an interesting yet bumpy path. Twenty-five years ago, a psycholinguist had only those intuitions to rely on if he or she wanted to consider the relationship between child language development in general and the development of idiomatic figurative competence in particular. Not until 1975 did Lodge and Leach publish, in JSHR, "Children's Acquisition of Idioms in the English Language", the first refereed article to present developmental evidence on children's comprehension of idioms in English. They reported that idiom competence was a late developing phenomenon, with 12 year-old children selecting a picture representing the figurative meaning of common idioms such as *spill the beans* and *kick the bucket* with about 50% accuracy. Because the understanding and use of figurative language was generally taken to be a skill developed in the period of formal operational thought, these results were not viewed, at least initially, as surprising.

Since 1975, research has added considerably to the developmental data base, describing a variety of stimulus and task variables that influence when and to what degree children can be judged to understand the figurative meanings of idioms. For example, a number of subsequent studies have shown that, in contrast with Lodge and Leach's results, children well below 12 years of age understand many common idioms (Ackerman, 1982; Cacciari & Levorato, 1989; Johnson, 1985; Prinz, 1983; ), and a majority of children understand many common idioms by 9 or 10 years of age (Brinton, Fujiki, & MacKey, 1985; Cacciari & Levorato, 1989; Gibbs, 1987, 1991; Johnson, 1985; Levorato &

Cacciari<sup>1</sup>, 1992). At the same time, the evidence does not support a straight line function of increasing figurative competence with increasing age (Abkarian, Jones & West, 1992).

There is similarly mixed evidence from research investigating the impact of various stimulus variables on the acquisition of figurative competence. An example of these mixed and/or conflicted findings can be found in the recent research focused on establishing the impact of an idiom's 'analyzability' (the degree to which the parts of an idiom contribute to its meaning; Gibbs, Nayak & Cutting, 1989; Cutting & Bock, 1997) on comprehension. Semantically analyzable idioms are reported to facilitate comprehension in some studies (Gibbs, 1987, 1991; Nippold & Rudzinski, 1993; Nippold & Taylor, 1995), but not in others (Abkarian, et al., 1992; Abrahamsen & Burke - Williams, 1996). Another important stimulus variable is familiarity – how frequently an idiom is encountered by some target population. More evidence shows that familiar idioms are easier to understand than unfamiliar idioms (Nippold & Rudzinski, 1993; Nippold & Taylor, 1995), but sometimes the effect of familiarity is not significant, particularly when the idiom is embedded within a rich story context (Levorato & Cacciari, 1992).

Certainly, the 'installed' knowledge base regarding the development of idiomatic competence has advanced substantially when compared to the initial efforts of 25 years ago. With equal certainty, there is more work to do. The intention of the present research is to do another piece of that work, by further examining issues relating to the 'analyzability' of idioms, and exploring whether the data on idiom acquisition developed using verb phrase idioms generalizes to other idiom forms.

### Objectives

The objectives of the research presented in this dissertation are three-fold. First, to reexamine the incompletely resolved issue of analyzability by using only unfamiliar idioms to probe the development of figurative comprehension. The objective is to add to the debate on analyzability data for which a principal confound— familiarity, has been effectively controlled. The

---

<sup>1</sup> All cited research for Cristina Cacciari and Maria C. Levorato used Italian idioms with native Italian speakers.

second objective is to investigate whether, and to what degree, the theoretical and developmental data on idiom comprehension, derived almost exclusively from verb phrase idioms, generalizes to other idiom forms. The issue is framed with reference to the impact of the syntactic form of idioms on children's developing comprehension competence. The goal of this portion of the investigation is to determine whether other idiom forms (noun phrase, adjective phrase and anomalous idioms) behave in the same manner developmentally as verb + noun phrase (verb phrase) idioms. As Nippold (1991) observes, the syntax of idioms ranges from single words (*nosey*) to independent clauses (*the cat's got his tongue*), and include verb, noun, adjective and adverb phrase idioms. Yet, verb phrase idioms have been used almost exclusively in developmental research to date (see, for example: Abkarian, et al.<sup>2</sup>, 1992; Ackerman, 1982; Ezell & Goldstein, 1991; Lodge & Leach, 1975; Prinz, 1983; and others). Given the real diversity of idioms, it seems necessary to establish whether, and to what degree, the data on verb phrase idioms generalizes to other idiom forms. The third objective will then be to determine whether semantically analyzable and unanalyzable idioms behave differently based on idiom form.

Before discussing this dissertation's treatment of analyzability, the generalization of findings from verb phrase idioms, and of the need to control for idiom familiarity in depth, certain 'foundation issues' are developed in the following section. These are framed as questions:

1) What is the origin of the current interest in idioms (see, for example, Cacciari & Tabossi, 1988; Everaert & van der Linden, 1989 (cf Everaert, et al., 1995); Everaert, et al., 1992, 1995, regarding levels of research interest)? This question is addressed in the section Historical Overview;

2) Where are idioms represented within the larger field of figurative language? This is addressed in the section Traditional Definitions of Figurative Language; and,

3) How can we define and differentiate an 'idiom' from figurative language in general? The functional definition for the term 'idiom' as used in this dissertation is presented in Idiom: A Working Definition.

---

<sup>2</sup> Nine of ten idioms.

These questions are presented as points which establish a logical basis for proceeding into a review of the research bearing on the topic areas of analyzability, generalizability, and familiarity. This review of the literature is presented in three sections:

4) The Framework, provides an organizing scheme for reviewing both theories of idiom processing and representation, as well as theories regarding the semantic flexibility of idioms;

5) Theories of Idiom Processing and Representation, where evidence from adult studies regarding how idioms are processed and where they are stored within the language system is discussed; and,

6) Studies of Typical Development, where the developmental evidence bearing on issues of analyzability, generalizability and familiarity is presented.

#### Historical Overview

Interest in studying how idiom comprehension, and how adults both process and mentally represent idioms has increased considerably over the last 25 years. This increase reflects both a shift in how idioms are treated by theories of language, and of a heightened awareness of the significance of figurative competence as a component of later language development.

Within linguistics during the 1960s and early 1970s, idioms were typically seen as either inconvenient stumbling blocks to theories of language (Chafe, 1968; Weinreich, 1969), or as a part of lexicography, and hence only incidentally related to linguistic theory (Everaert, van der Linden, Schenk, & Schreuder, 1995). Generative grammarians (Chomsky, 1980; Fraser, 1970; Heringer, 1976) and researchers interested in natural language processing (Becker, 1975; Gasser & Dyer, 1986) treated idioms as noncompositional or semantically unanalyzable word strings. In this view, an idiom's meaning was viewed as an arbitrary and unitary entry into a mental lexicon. So defined, there could be no analyzable (or transparent) relationship between the parts of an idiom and its figurative meaning, in the same way that the relationship between any single word and the object it represents is completely unanalyzable and opaque. With idiom competence thus conceptualized as some sort of word-list learning exercise, the study of idioms did not generate much early research

interest.

But why, in the first place, this view of idioms as opaque and unanalyzable lexical items? After all, even casual reflection shows that idioms can be composed of many words (for example, *give him enough rope and he'll hang himself*, or *fly by the seat of your pants*), with relatively complex internal structure. Treating idiom meanings as opaque and lexicalized came about in part because most theories of meaning were built, per Frege (1879/1972), on the notion of compositional analysis. In a compositional analysis, sentences are treated as inputs into a logical calculus, with sentence meanings derived by a rule-based concatenation of the meanings of the component words. Yet idioms and other figures of speech appear to defy a compositional analysis. For example, the figurative meaning of the idiom *he let the cat out of the bag* cannot be paraphrased as 'he started some trouble' by analyzing the literal meanings of the individual words of the idiom (Hoffman & Honeck, 1980). Something other than a literal, compositional analysis of the words comprising the idiom is required to understand the figurative meaning. Since idioms resist the application of a logical calculus, taking them *off the boards* (by defining them as noncompositional strings) seemed to resolve the challenge of accounting for them within formal theories of language. However, several lines of evidence and reasoning suggest this action was premature, and these will be discussed shortly. First though, to *get a handle* on what idioms are and are not, to ensure all readers *in the same ballpark*, definitional issues are considered.

### Traditional Definitions of Figurative Language

Honeck's (1980) comment that "no general, commonly accepted criteria exist by which figurative language phenomena can be distinguished from one another or from non-figurative phenomena" is still largely true. However, some means for differentiating among the various categories of figurative forms need to be stated. This will be done by suggesting criteria useful for categorizing various figurative forms, and providing examples of each. Imbued with a Roschian spirit, perhaps there can be agreement on where the robins go, if not the penguins.

The common characteristic of the figurative forms defined below is that

there are differences between the meaning of the words comprising their surface forms and the underlying figurative meanings. To paraphrase our friends in computing, 'what you see is not what you get'. When we say that *the Sonics are a hot team*, we mean to say that the Sonics are on a winning streak, not that they possess abnormal metabolisms (excepting perhaps the former Sonic Shawn Kemp). As Cacciari and Levorato (1989) note, this differentiation between the literal and figurative meanings separates these forms from other types of figurative language, such as irony and sarcasm, in which "the meaning is different from the communicative intention" (p. 388). This differentiation also partially separates these figurative forms from other non-literal uses of language such as indirect requests, in which conventionalized forms are used to convey specific communicative intents in particular contexts.

Idioms. Figurative utterances are conceptualized as extending 'along a gradient of originality' (Estill & Kemper, 1982; see also Burbules, Schraw, & Trathen, 1989) or 'productivity' (Clausner & Croft, 1997). Along this gradient, or axis, idioms are held to be the least original or generative of figurative forms, and novel metaphors the most original and generative. Idioms are least original in that they are lexically fixed word strings. The figurative meanings are assigned by convention and are rigidly related to each idiom (Ackerman, 1982). For example, to *have a cow* means to 'be really upset' by convention, not by compositional analysis. Analyzing the individual words comprising the idiom does not help the listener to derive the idiomatic meaning (Seidl & McMordie, 1978). Also, the name of another large mammal cannot be substituted for cow without affecting the idiom's meaning: 'to have a horse' does not convey the same figurative meaning. So in this sense, the lexical string is inflexible. These broad characteristics reflect and support the general view of idioms as noncompositional, lexical strings. They are 'complex units' (Everaert, et al., 1995); complex in their being multi-word compositions, units in referring to individual underlying meanings. However, the points offered here as defining idioms do overlap to greater and lesser degrees with the definitions of other figurative forms. To further clarify what idioms 'are', the definitions of other common figurative forms are presented next, to show what idioms 'are not'.

Metaphors. Contrasting with the conventionality of idioms is the

productivity and creativity associated with metaphors. Metaphors are a figure of speech in which the thing commented on, the 'topic', is compared to another, the 'vehicle', on the basis of one or more common semantic features, the 'ground' (Richards, 1936; Perrine, 1971). For example, to quote a metaphor of metaphor by Paivio (1979), "for the student of language and thought, metaphor is a solar eclipse". Here the topic is 'metaphor' and the vehicle 'solar eclipse'. The 'ground' is, as Paivio notes, that a metaphor "hides the object of study (linguistic meaning) and at the same time reveals some of its characteristics (semantic creativity)" (p.151) in the same way the moon hides the sun, yet facilitates the study of its corona. Since the number of vehicles and topics is essentially infinite, metaphors are obviously generative and novel. Recent views of metaphoric production and use stress that metaphors are not exceptional or unusual forms of language, but rather are routine parts of everyday conversation (Glucksberg, 1989; Lakoff, 1987; Lakoff & Johnson, 1980).

Proverbs and similes. In a middle ground between the conventionality of idioms and the productivity of novel metaphors are proverbs and similes. Proverbs are wise sayings which, like idioms, have meanings conventionalized over time in specific cultural settings. Unlike idioms, however, a sense of the underlying meaning can usually be derived from a broad reading of the words comprising the proverb— 'a penny saved is a penny earned', 'a stitch in time saves nine', and so on. Similes are limited case variants of metaphor. Similes make the comparison between the topic and vehicle more explicit by using the word 'like' or the phrase 'as + adjective + as' (as crazy as, as big as, etc.) between the topic and vehicle.

#### Idiom: A working definition

What, then, is an idiom? A synthesized majority view of what an idiom is would read something like this: An idiom is a complex unit. It is composed of a conventionalized word string that is rigidly associated with a specific figurative meaning. However, the idiomatic meaning cannot be derived directly from an analysis of the literal meanings of the words comprising the idiom. Further, an idiom cannot be paraphrased freely and still retain its conventionalized

meaning. Finally, an idiom's figurative meaning is relatively independent of context. This is presented as a working definition, not as an 'absolute'. As will be seen in the next section, there are grounds for modifying this definition. For now, let this definition stand as a 'benchmark'— a generally agreed on reference point.

When one uses a benchmark, measurements of change and progress can be made with some reliability. In that spirit the working definition defines a point for moving into the review of the literature, and this is presented next.

### Review Of The Literature

This section is divided into three segments: (1) Framework; (2) Theories of Idiom Processing and Representation; and (3) Studies of Typical Development. Before moving ahead, I want to first present a few advance organizers. The intent of the 'Framework' is self-explanatory. Current psycholinguistic theory and research involving idioms can be categorized into 3 - 6 primary channels<sup>3</sup>, with a certain amount of polysemy, synonymy and theoretical inter-twining. The Framework provides some guidance for navigating these tricky waters. Next, the studies presented in 'Theories of Idiom Processing and Representation' focus on identifying the task and stimulus variables which influence how adults process idioms. The adult studies are reviewed before the developmental studies because: i) much of the developmental research arises from hypotheses initially investigated with adults, making interpretation of the developmental evidence problematic without this theoretical framework; and ii) there is a need to specify the end-points of typical development. That is, we can make most sense of the developmental evidence if we first identify how idioms are understood by mature users of the language, and how idiomatic meanings might be stored in the lexicon.

Finally, the works presented in 'Studies of Typical Development' focus on answering the questions 'when do children understand the figurative meanings of idioms?' and, more recently, on identifying 'what task and stimulus variables

<sup>3</sup> The term 'channel' is used deliberately. Some of the theoretical positions on idioms are strongly segregated, forming different 'rivers' of thought. Others, however, segregate on some issues, then recombine into a common stream of thought. They behave like channels of a large river, moving apart, then back together.

influence children's comprehension of idioms?'. The proximate goal in reviewing this literature is to construct a developmental profile of idiom comprehension, then to identify how various task and response conditions impact children's ability to demonstrate figurative competence. Ultimate goals include identifying how children learn and process the figurative meanings of idioms, and determining whether processing strategies change over the course of development. The unifying goal for presenting these adult and developmental studies is to aid the reader in integrating the psycholinguistic evidence on idioms into more general theories of language competence.

### Framework

Broadly, the major topic areas in idiom research can be organized along three major and one minor axes. The major axes are: (1) How are idioms processed?; (2) How are idioms mentally represented?; and, (3) How do idioms behave internally -- their syntactic and (central to this dissertation) semantic flexibility? (Everaert, et al., 1995; Forrester, 1994; Keysar & Bly, 1995; Schweigert, 1991; Schweigert & Moates, 1988). The minor axis is: How does an idiom's familiarity impact its comprehension (Forrester, 1994; Kemper, 1986; Keysar & Bly, 1995; McGlone, Glucksberg, & Cacciari, 1994; Schweigert, 1986, 1991)? Because the processing of a meaningful complex unit demands a 'look-up' stage, there are at least one-way interactions between the processing and representation axes. Thus, information on these axes is presented together, and includes information on: a) the processing model; b) how idioms are represented or stored, given the particular processing model; and c) the theoretical perspective underpinning each position, with selected references. This information (Process - Representation - Theory/References) is presented next, schematically, and is keyed and indented in the following sections for readability, as follows:

#### **Process-**

#### Representation

#### Theory/References

### Process and Representation

**Literal First** - When idioms are encountered, "the literal meaning of an idiom is processed and evaluated for appropriateness before the figurative meaning" (Schweigert, 1991, p.305). 'Multiple-processing' of idioms is required in order to recover an idiom's figurative meaning.

Idioms are stored in an idiom list that is separate from the general lexicon (Newmeyer, 1974; Weinreich, 1969).

This model is identified with the traditional view of idioms as lexically fixed exceptions to the canon of compositionality

(Katz & Postal, 1963). Meaning is not derived by a rule based concatenation of the meanings of the words that make up the idioms. Instead, one conventional, figurative meaning is directly associated with the idiom as a unit. Positive evidence supporting this theoretical position was reported by Bobrow and Bell (1973). The literal first model would predict that literal phrases are processed more rapidly than idiomatic phrases. This would be so because idioms require a second 'reprocessing' stage when the attempted literal gloss does not fit the context. Unfortunately for the model, as Schweigert (1991) notes, the processing time advantages which should be seen in reaction time experiments for literal phrases, haven't been found .

**Lexical Representation** - When idioms are encountered, both the literal and figurative meanings are activated simultaneously; also called the 'simultaneous processing' model.

Idioms are stored in the general lexicon as single giant lexical chunks.

This model, while not strictly identified with the traditional view of idioms, does postulate that idioms are giant noncompositional chunks. Swinney & Cutler (1979) developed this model. They showed in reaction time experiments that, contrary to the predicted slowed processing times for idioms in the **literal first** model (due

to literal first, then subsequent figurative look-up), idioms were processed consistently faster than literal control strings.

**Figurative First** - When idioms are encountered, the figurative meaning is activated first. The literal meaning may never be activated. Also called the 'idiomatic processing' (Forrester, 1994; Schweigert, 1991) or 'direct access' (Reagan, 1987) model.

Idioms are stored in the general lexicon, but may be partially analyzable or compositional.

This model is supported by a variety of reaction time studies, as summarized in Schweigert (1991)<sup>4</sup>, but has been challenged by findings that do not show processing time differences between the literal and figurative uses of idioms (where both interpretations are possible - *kick the bucket, lose your marbles*, etc.; see Estill & Kemper, 1982; Schweigert, 1986). This model is best, though not strictly, associated with theories of idiom 'decomposability' (Gibbs & Nayak, 1989; Nunberg, 1978; Wasow, Sag & Nunberg, 1983).

The debates surrounding these theories of idiom processing and mental representation have not yet been resolved. This is in part because replication and extension studies have not produced consistent outcomes (Schweigert, 1991). Indeed, the lack of consistency in research outcomes led to a more detailed reexamination of the stimuli – that is, the idioms, themselves. This was due to a realization among researchers that the presumption of homogeneity of idioms was in error in important ways.

### Flexibility

The third major axis in idiom research addresses the flexibility of idioms. There are three main streams of thought regarding the flexibility of idioms.

These three areas are outlined as: (a) hypotheses; with, (b) terminology

<sup>4</sup> Gibbs, 1980, 1986; Mueller & Gibbs, 1987; Ortony, Schallert, Reynolds & Antos, 1978; Schweigert & Moates, 1988.

associated with each hypothesis from the literature; and, (c) the theoretical perspective underpinning each position. This information (Hypothesis - Terminology - Theory/References) is again presented schematically, and is keyed and indented as follows:

## **Hypothesis**

### Terminology

### Theory/References

**Idioms are inflexible** -- Idioms are units, they are learned as whole chunks, and they cannot be altered.

Lexical unit, or chunk; giant word (string); lexicalized; frozen; dead metaphor.

This hypothesis is closely associated with Swinney & Cutler's (1979) Lexical Representation Hypothesis, in which idioms are stored in the general lexicon as giant unitary word strings (see also Fraser, 1974, cf Swinney & Cutler, 1979; and Heringer, 1976).

**Idioms are syntactically flexible** -- Idioms are expressions which can, in principle, "undergo any syntactic operation their literal counterparts can undergo" (Schenk, 1995, p. 254). Syntactic flexibility does not impact meaning, which remains an inflexible, single semantic unit.

Frozen; dead metaphor; frozenness hierarchy.

Summarizing from Everaert, et al. (1995) there is general agreement that the syntax of phrasal idioms (Fraser, 1970) should behave more or less like the syntax of literal phrases (see also Newmeyer, 1974; Jackendorf, 1975 for additional discussion). However, cumulative study has shown that idioms vary considerably in the kind and degree of syntactic flexibility they exhibit (Gibbs & Nayak, 1989; Gibbs & Gonzales, 1985), while retaining their figurative meaning. It was not initially clear theoretically why this was so. One mechanism proposed to account for the differences in

flexibility was that idioms were in fact semantically flexible.

**Idioms are semantically flexible** – Idioms are not single units. Rather, they are complex units, and an idiom's parts can contribute to the overall figurative meaning (Cutting & Bock, 1997; Gibbs & Nayak, 1989; Gibbs, Nayak & Cutting, 1989; Nunberg, 1978; Wasow, et al., 1983). Semantic flexibility varies along a continuum, from idioms that are analyzable/decomposable to idioms that are unanalyzable/nondecomposable.

Analyzable/decomposable idioms (e.g.: *blow off some steam*): The parts of an idiom contribute to its overall meaning. An analysis of the units making up a semantically analyzable idiom aids in extracting the figurative meaning. Analyzable idioms are typically metaphorically transparent. That is, "the expression *keep a straight face* is transparent because the nonliteral meaning, to show no emotion, is closely related to the literal meaning" (Nippold & Taylor, 1995, p.427). At the opposite end of the analyzable - unanalyzable continuum are idioms that are:

Unanalyzable/nondecomposable (e.g.: *throw in the towel*). The parts of the idiom do not contribute to the overall meaning. An analysis of the units making up an unanalyzable idiom does not aid in extracting the figurative meaning. Unanalyzable idioms are metaphorically opaque. The idiom *talk through one's hat* is opaque because the nonliteral meaning of this idiom, to not know the facts, has little to do with the literal meaning" (Nippold & Taylor, 1995, p.427).

Closely associated with Gibbs 'semantic decomposition hypothesis' (Gibbs & Nayak, 1989; Gibbs, Nayak & Cutting, 1989), and more generally with theories of 'conceptual metaphors' (Lakoff, 1987; Lakoff & Johnson, 1980). The argument that idioms are semantically flexible runs strongly counter to traditional views of idioms as single inflexible semantic items. If idioms 'decompose', then in principle they are following the canon of compositionality. If idioms decompose, that is if parts of meaning can be pulled apart

from the figurative meaning, and these parts can be shown to contribute independently to the overall figurative meaning, then idioms are not exceptions to more general language processes.

The research relevant to the axes introduced in the Framework -- how idioms are processed, mentally represented, and their internal behavior (flexibility) -- is summarized in Table 1, and reviewed in Theories of Idiom Processing and Representation next.

Table 1

Theories of Idiom Processing, Lexical Representation and Idiomatic Flexibility

Theory	Processing	Representation	Flexibility
Literal First	Literal, followed by figurative if needed.	Idiom list; as lexical units.	Inflexible, or flexible syntax. Unanalyzable.
Lexical Representation	Literal and figurative are simultaneous.	General lexicon; as lexical units.	Inflexible Unanalyzable
Figurative First	Figurative, followed by literal if needed.	Not specified	Flexible syntax and semantics. Analyzable

Theories of Idiom Processing and Representation

The Literal-First Hypothesis: Idioms in their own list

The literal - first hypothesis argues that the "literal meaning of an idiom is processed and evaluated for appropriateness before the figurative meaning" (Schweigert, 1991). Research from Bobrow & Bell's (1973) investigations of idioms supports this hypothesis, and is reviewed next.

Bobrow and Bell (1973) hypothesized that when an idiom is encountered

in an appropriate context, a special 'idiom processing mode' is activated. This processing mode somehow allows the figurative meaning to be accessed directly in a special idiom lexicon, in which the idiomatic phrases are stored as single lexical items. This activation apparently occurs after the listener has attempted to first process the idiom literally, or at least until some threshold is reached that rejects the literal analysis in favor of the idiom processing mode. The meanings of literal phrases are, on the other hand, constructed from the separate meanings of the words "mapped into a semantic whole" (p. 343).

To test their reasoning empirically, Bobrow and Bell asked adults to report whether they saw either the literal or idiomatic meaning of ambiguous sentences first under one of two priming 'set' conditions (the set consisted of reading and interpreting four literal or four idiomatic sentences before encountering the target sentence). The responses were compared to a no-set baseline. Results showed that subjects were significantly more likely to see an idiom's figurative meaning first in the idiom-set condition, and less likely to see a figurative meaning first in the literal-set condition. Bobrow and Bell argued that these results support the idea that there are distinct processing modes used when literal and idiomatic ambiguities are encountered, with the literal mode employed first, and the idiomatic mode engaged only as the attempted literal interpretation is unsuccessful.

Bobrow and Bell's theorizing would predict, among other things, that it should take longer to process idioms than nonidiom control strings (Swinney & Cutler, 1979). As will be seen next, this prediction has not been borne out by a number of reaction time studies.

### The Lexical Representation Hypothesis: Idioms as giant lexical strings

The Lexical representation hypothesis, as developed by Swinney and Cutler (1979), has two main arguments: First, that idioms are stored in the general lexicon as giant word strings; second, that both the literal and figurative meanings of idioms are accessed simultaneously and directly. Their view that idioms are stored in the general lexicon differs from some aspects of the traditional treatment of idioms, while the hypothesis that meanings are treated as single lexical units generally reflects the traditional view of idioms in the

literature (Keysar & Bly, 1995).

The evidence Swinney and Cutler used to develop their hypothesis comes from a series of reaction time (RT) experiments. Participants were required to judge idioms and nonidiom control strings for overall 'sensibleness' or grammaticality. Swinney and Cutler reasoned that, because the task required each word string to be analyzed as a unit, participants would show longer RTs for grammatical idioms than for nonidiom controls if idioms had to be interpreted through recourse to a special 'idiom processing mode', per Bobrow and Bell (1973). If, on the other hand, idioms are stored as unitary lexical entries, RTs for idioms should be faster than RTs for literal control strings. This would be so, the authors argued, because the computation of both literal and idiomatic meanings would be done simultaneously when the idiom string-as-one-lexical-entry was encountered. This should be accomplished more rapidly than the analyses of the various relationships among the lexical items in the nonidiomatic strings.

The results supported their hypothesis, with RTs significantly faster for the idiom than for the non-idiom strings, refuting the idea that a special idiom processing mode was activated, and supporting the Lexical Representation Hypothesis.

However, several cautions are in order for interpreting these results. First, while the RT evidence does support the thesis that idioms are stored as giant lexical strings, there is no direct evidence supporting the simultaneous activation of literal and figurative processing. Indeed, such parallel processing would only occur when idioms with both well-formed literal and idiomatic meanings were processed. Yet, these idioms are only a subset of all idioms. Simultaneous processing may not be activated for either literally ill-formed (*hold your peace*) or anomalous idioms (*by and large*), for which there are no literal glosses. From the data, we simply do not know if simultaneous activation would occur with either anomalous or ill-formed idioms, or not. Indeed, there are very little data regarding the behavior of anomalous idioms reported in the literature, and it is one purpose of this dissertation to investigate these 'oddball' or anomalous idioms directly, to begin to fill in this gap in our knowledge.

An additional caution in interpreting Swinney and Cutler's results focuses on the high likelihood that idioms used idiomatically are encountered

far more often than are their literal paraphrases (Glass, 1983; Popiel & McRae, 1988). That is, while we might encounter *spill the beans* literally, it is much more likely that we will encounter it idiomatically. This frequency/familiarity effect may bias RT outcome studies of idioms, or at least of those studies that use context to bias a listener toward either a literal or figurative gloss. Indeed, as Schweigert notes, "The familiarity of the idioms and the context in which they have appeared have not been controlled for in much of the previous research and thus might account for some of the confusion in the literature" (Schweigert & Moates, 1988, p.283). The potential confound of familiarity/frequency will be discussed again later, in relation to differences in the degree of the familiarity of idioms themselves.

Subsequent investigations using RT paradigms have generally supported Swinney and Cutler's contention that idioms can be stored as giant word strings, based on the more rapid processing of these idioms in RT experiments, when compared to non-idiomatic controls (Cronk & Schweigert, 1992; Gibbs, 1980; Gibbs & Gonzalez, 1985; Gibbs, Nayak, & Cutting, 1989; Glass, 1983). However, the claim that both an idiom's literal and figurative meanings are retrieved simultaneously has been challenged by some of these investigations. It is to this evidence that we now turn.

### Figurative First

Studies by Gibbs (1980, 1986; see also Mueller & Gibbs, 1987) suggest that the figurative meaning of idioms may be accessed directly, with no required access to the literal meaning. Gibbs (1980) used a sentence verification RT task to probe the influence of context on the interpretation of idioms with both literal and figurative meanings. Short stories biased the reader toward either the literal or figurative meaning of the idiom. Then, participants judged whether either a literal or figurative paraphrase of the target idiom meant the same thing as the idiom. Results showed faster RTs for the idiomatic meanings in the idiom-bias contexts, supporting the Lexical Representation Hypothesis.

However, in a follow-up recall task participants were given prompts of the idiomatic meanings of idioms used in both the literal-bias and idiom-bias conditions. For example, in the sentence verification task a participant might

hear a literal-bias story for *he's singing a different tune* (i.e., 'Jackson Browne's singing a different tune from the one we heard on the radio'), but in the recall task the prompt would cue the idiomatic 'changed mind', not the literal 'singing a different song'. Here, participants prompted recall of the idioms used in the literal-bias condition was better than their prompted recall of the idioms used in the idiom-bias condition. Hoffman and Kemper (1987) argue that this is because the nonliteral meanings may be accessed directly, without any requirement to construct the literal gloss. That is, recall was enhanced in the literal-bias condition because the participants are predisposed to access the figurative meanings of idioms first, and have to work harder in order to gloss the alternate, literal interpretation. Thus, while idioms may be lexically represented as giant single words, they do not necessarily undergo simultaneous processing to access their figurative meaning.

Several other studies have also found processing time advantages for idioms used figuratively over matched literal controls (Ortony, Schallert, Reynolds, & Antos, 1978; Schweigert & Moates, 1988). However, the findings for a processing time advantage are not supported uniformly. A reaction time study by Estill and Kemper (1982), did not find a significant processing time difference between idioms used figuratively (e.g.: "By the fourth day in the hospital, Orville was *climbing the walls* to go home") and the same idioms used literally (e.g.: "Orville was interested in spiders and could sit for hours and watch them climbing the walls of his garden"), as would be expected from 'figurative first' processing. Estill and Kemper (1982) interpreted their results as supporting simultaneous processing modes (Swinney & Cutler, 1979). Similar outcomes -- no processing time advantages for idioms used figuratively versus idioms used literally -- were reported by Schweigert (1986). These data, by countering predictions derived from 'figurative first' processing, suggest that figurative first processing does not completely explain the figurative processes related to idiom comprehension.

So, in a question first raised in the Framework, above, what sense can be made of these conflicting outcomes from studies interested in the same linguistic phenomena, and often sharing the same (or at least similar) research design? Part of an answer lay in the growing realization that idioms are not

homogeneous in important ways -- that they differ in important characteristics - particularly in their flexibility, and in their familiarity.

### Characteristics of Idioms

#### Flexibility.

The construct of an idiom's 'flexibility' refers to two somewhat different aspects of the linguistic behavior of idioms - syntactic and semantic flexibility. Syntactic flexibility refers to the ability of some (but not all) idioms to undergo various syntactic transformations while retaining their idiomatic meaning (see Newmeyer, 1972, 1974; Jackendorf, 1977; and Schenk, 1995<sup>5</sup> for additional discussions on the variety of syntactic operations possible). As noted in the 'Framework', above, the demonstration of syntactic flexibility does not significantly challenge traditional views of idiomatic meaning, in which the parts of an idiom represent a single semantic unit (i.e., the idiom is still noncompositional).

Semantic flexibility, on the other hand, refers to the thesis that the parts of an idiom contribute compositionally to its overall figurative meaning (Cutting & Bock, 1997; Gibbs & Nayak, 1989; Gibbs, Nayak & Cutting, 1989; Nunberg, 1978; Wasow, et al., 1983). The claim that idioms are semantically flexible is a strong one, in that this claim challenges the widely held view that idioms are rigid, fixed and noncompositional. If idioms 'decompose', then they behave semantically like novel, generative utterances. The thesis that at least some idioms are semantically flexible is closely associated with the work of Gibbs and colleagues (Gibbs & Gonzales, 1985; Gibbs & Nayak, 1989; Gibbs, Nayak, Bolton & Keppel, 1989).

As has already been discussed, the research does not speak with one voice regarding whether reaction time advantages do or do not accrue to

<sup>5</sup> Schenk (1995, p.254) provides examples of some of syntactic operations an idiom can in principle undergo:

#### transformed but figurative idiom

- a. It was Mary's leg that Pete pulled.
- b. At whose door did Pete lay his failure?
- c. Whose leg did Pete say that Mary pulled?
- d. Mary's heart Pete broke.
- e. The beans seem to have been spilled by John.
- f. The roof seems to have caved in  
on John's dream.

#### operation

- clefted, direct object as focus  
Wh-movement, PP *at whose door*  
Wh-movement, NP *whose leg*  
topicalize NP *Mary's heart*  
passivize idiom, then raise NP *the beans*  
raising of NP *the roof*

idioms. In addition, idioms do vary in the degree to which they can be altered syntactically while retaining their figurative meaning (Schenk, 1995; Gibbs & Gonzales, 1985; Gibbs & Nayak, 1989). Gibbs and colleagues hypothesized that one way to account for these discrepancies would be to consider that idioms are not unitary semantic units, sometimes syntactically flexible but always semantically unanalyzable. Rather, they proposed that: i) idioms are semantically analyzable (Gibbs & Nayak, 1989; Gibbs, Nayak, Bolton & Keppel, 1989; Gibbs, Nayak, & Cutting, 1989); and ii) the analyzability of idioms varied along a continuum. Gibbs argues that it is these factors – that idioms are analyzable, and that analyzability exists along a continuum – which account for both the differences in reaction time research outcomes, and for the differences in the syntactic flexibility of idioms. These propositions are discussed next.

Gibbs' experimental investigations into the semantic behavior of idioms derived from a line of reasoning first proposed by Nunberg (1978). Nunberg argued that "idioms do not form a unique class of linguistic items, perhaps 'dead metaphors', but share many of the compositional properties normally associated with more literal types of speech" (cf Gibbs, Nayak, Bolton & Keppel, 1989, p.66). In this argument, idioms are not so 'idiomatic' as linguists' intuitions would lead one to believe.

In discussing the analyzability of idioms, Gibbs, Nayak, and Cutting (1989) are careful to first note what semantic analyzability/decomposition is not: It is not equivalent to 'lexical decomposition', where meaning is derived by analyzing each lexical item into its "semantic primitives" (p. 577). Instead, the decomposition of idiomatic meanings is a rather flexible process, wherein the parts of an idiom can contribute either literal or figurative meanings toward the overall interpretation of the whole idiom. Given this somewhat 'fuzzy logic' approach to defining analyzability, I will let Gibbs describe what is meant by semantic analyzability/decomposition:

One way we have characterized idiom decomposition is in terms of *semantic fields* which are clusters of lexical items that share some nuclear concept or feature (Miller & Johnson-Laird, 1976) or which go together because they pertain to the same knowledge about some event (Kittay & Lehrer, 1981), or to the same "experiential gestalt" (Lakoff & Johnson, 1980). Idioms are more analyzable to the extent that their individual components share the same semantic fields with their

idiomatic references. For example, the individual parts of *pop the question* must be in the same semantic field, or conceptual domain, as their idiomatic references "propose" and "marriage" for this idiom to be viewed as decomposable. However, the individual components of phrases such as *kick the bucket* or *chew the fat* are not in the same semantic fields as their respective figurative referents (i.e., "to die" and "to talk without purpose") and should not be viewed as semantically decomposable... [and] it is difficult for people to assign the individual parts of these idioms with particular parts of their overall figurative meanings (Gibbs, Nayak & Cutting, 1989, p.578).

This would seem to make the process of categorizing idioms as 'Analyzable' or 'Unanalyzable' somewhat dicey. Indeed, as Gibbs notes, "there are no well-defined procedures for specifying whether a given idiom is semantically decomposable or not" (Gibbs, et al., 1989, p.578). This observation is accepted at face value, and will have a direct impact on the Experimental Procedures used in this dissertation. Functionally, it means that for any investigation into the construct of semantic analyzability, considerable procedural description and control must be exercised if experimental results hope to generalize beyond the bounds of a particular investigation. The specific procedures used in this investigation are contained in Methods (Stimuli - Analyzability Ratings), and in Appendix A. Since the procedures in Appendix A are adapted from Gibbs' methodology, his original procedures are discussed next.

Before investigating semantic analyzability, Gibbs (Gibbs & Nayak, 1989) first needed to determine if adults could make reliable judgments about whether the individual words in verb phrase (VP) idioms contributed to each idiom's figurative meaning. To test for this, students first divided a mixed list of normally analyzable<sup>9</sup> (*miss the boat*), abnormally analyzable (*pass the buck*), and unanalyzable (*chew the fat*) idioms into two parts. This was done by:

- a) comparing each idiom to its literal paraphrase, and then;
- b) deciding for which idioms the individual components contributed to their overall figurative meanings (analyzable idioms), and which idioms the individual words did not make such a contribution (unanalyzable);
- c) next, the same participants divided the analyzable list into two parts,

---

<sup>9</sup> The idioms in each category were initially selected based on the authors' intuitions.

one where the words of an idiom had literal relations to the figurative referents (normally analyzable), the second composed of idioms whose words had a "metaphorical" relation to their meanings (abnormally analyzable).

Gibbs and Nayak found that adults could make reliable judgments about whether the individual words in verb phrase idioms contributed to each idiom's figurative meaning. The mean proportion of agreement among participants for each categorical decision was: 79% for abnormally analyzable idioms, 86% for normally analyzable idioms, and 88% for unanalyzable idioms (replicated in Gibbs, Nayak, Bolton & Keppel, 1989). Gibbs and Nayak argued from their results that: (1) idioms do differ in their degree of semantic analyzability; and, (2) adults can make statistically reliable classification decisions based on these differences.

To obtain more direct, on-line evidence regarding semantic analyzability, Gibbs, Nayak & Cutting (1989) asked adults in an RT paradigm to make judgments about the meanings of various English phrases. The idiomatic phrases were either normally decomposable (analyzable), abnormally decomposable, or nondecomposable (unanalyzable) idioms (from Gibbs & Nayak, 1989) These were matched with nonidiomatic grammatical control strings and with ungrammatical, nonidiomatic word strings. Gibbs et al argued the following:

1) all idiom RTs would be faster than the RTs of their literal control strings, because there would be no need to access and add up the literal meanings first (per Swinney and Cutler, 1979; Gibbs & Gonzalez, 1985);

2) abnormally analyzable and unanalyzable idiom RTs would be faster than the RTs for normally analyzable idioms. This is because the unanalyzable (most lexicalized) idioms would require relatively little compositional analysis, while the normally analyzable idioms would require more (if idioms were routinely interpreted via a compositional parsing strategy).

The results only partially supported the original predictions:

1) normally and abnormally analyzable idiom RTs were significantly faster than their literal controls strings; however,

2) unanalyzable idioms took longer to verify than both the control strings and the other categories of idioms, contradicting Gibbs and Gonzales, 1985

results, and what would be expected if unanalyzable idioms were highly lexical. Gibbs reasoned that this was because it was harder to "assign independent meanings to the unanalyzable idioms' individual parts" (p. 583). RTs are slower apparently because some sort of additional processing is required. On the other hand, ordinary compositional processing is used to derive the meanings of both analyzable idioms and literal phrases. No other type of processing mode is needed to analyze the components making up the analyzable idiom's meaning. Analyzable idiom RTs are nevertheless faster than for the (novel) matched literal control strings because idioms are still conventionalized, albeit modestly flexible. For unanalyzable idioms, on the other hand, a compositional analysis is not possible, and 'something new' has to be done – a look up of the meaning directly stipulated in the mental lexicon. This, apparently, takes longer to do because a compositional analysis has to first be tried, then rejected, before the idiom is 'looked up'<sup>7</sup>.

Gibbs' research led him and his colleagues to relate the observed semantic flexibility of idioms to a theoretical model of idiom processing -- semantic decomposition – which allows for the empirical testing of predictions about the behavior of idioms, based upon the model. However, while the predicted processing time advantages for normally analyzable idioms were seen, the predicted processing times for unanalyzable idioms were not verified, indicating that this theoretical model was incomplete.

Indeed, several other studies have also reported outcomes that do not completely support the semantic decomposition model. For example, Cutting and Bock (1997), in a cued-recall study of idiom production, found no main effect for analyzability. In two developmental studies (Abkarian, et al., 1992; Abrahamsen & Burke-Williams, 1996, reviewed in detail in 'Studies of Typical Development'), no significant differences were found in children's comprehension of transparent and opaque idioms, indicating that transparency per se was not influencing acquisition processes.

Certainly, the outcomes of research on the construct of semantic analyzability do not speak with uniformity. However there is a general sense

<sup>7</sup> Though 'where' the look - up occurs is not clear; as van de Voort & Vonk (1995) observe, "The Decomposition Hypothesis has one major shortcoming: It is not clear how and where the meanings of the components of ...decomposable idioms are stored and how these meanings are accessed." (p.285).

that some idioms are more 'transparent' than others. There is, intuitively, a more direct link between *skating on thin ice* and 'danger', than between *kick the bucket* and 'to die' - and anecdotally, many children do seem to use a decomposition strategy in trying to understand idioms at some point in language development. Thus, we will want to first reexamine whether other stimulus variables influence idiom comprehension before we *throw out the baby with the bath water* in disregarding the construct of the semantic analyzability of idioms. Recent research has shown that the familiarity of idioms impacts the processing of idioms, and this area is examined next, as the fourth axis of research on idiom comprehension.

### Familiarity.

The impact of familiarity on idiom comprehension by adults would seem to be self-evident: The more familiar the idiom, the faster and more accurate the identification or retrieval of the figurative meaning, on whatever performance measure is used, with the effects of familiarity extending along a continuum. Indeed, 'familiar = fast' fairly summarizes the primary investigations to date encompassing familiarity (Forrester, 1995; Kemper, 1986; Keysar & Bly, 1995; McGlone, Glucksberg, & Cacciari, 1994; Schweigert, 1986, 1991; but see also Schraw, Trathen, Reynolds, & Lapan, 1988). However, such general statements about familiarity do not capture some subtle, but potentially important, interaction effects between familiarity and performance. For example, at the opposite end of the continuum of familiarity, unfamiliar idioms sometimes behave like nonidioms on various measures of performance (Forrester, 1995). Some of these subtle effects (for adults) may be more powerful effects for children who are in the process of acquiring of idiom comprehension. In order to *cover all the bases*, these subtle effects are detailed below.

A study by Forrester (1995) raises questions about 'just what is familiar?' when talking about an idiom's familiarity. Forrester investigated the influence of, and interactions among, idiom familiarity (familiar vs unfamiliar idioms), context (figurative vs literal bias), and canonical form (idioms in their standard, canonical form vs as semantically equivalent paraphrases) on reading time. As expected, a significant main effect for familiarity was found, with familiar idioms in both figurative and literal bias contexts read more rapidly than unfamiliar

idioms. However, an unexpected outcome, and one which was interpreted as supporting Gibbs' semantic decomposition hypotheses, was also found. There were no significant reading speed differences between the familiar idioms in their canonical form and reading the semantic equivalents of familiar idioms. And, this was true in both the figurative bias and literal bias story contexts.

Forrester interpreted these findings as indicating that it is the access to a familiar idiom's conventionalized figurative meaning (not to its surface form) which provides the reading speed advantage. This is because participants read both the canonical idioms and their 'semantically equivalent' phrases at equivalent speeds. Apparently, what is 'familiar' about a familiar idiom is the accessed figurative meaning, not the idiom's canonical form per se. Hence, Forrester's argument is generally compatible with the semantic decomposition hypothesis, or at least with Gibbs' argument for interpreting idioms via conceptual metaphors (Gibbs, Bogdanovich, Sykes, & Barr, 1997).

Keysar and Bly (1995) suggest another consideration when studying the impact of familiarity on idiom comprehension. They hypothesize that there is an interaction effect between an idiom's familiarity and its perceived transparency. Specifically, they argue that:

- i) use strengthens the connections between a complex conventional expression such as an idiom, and its conventional, stipulated meaning; and,
- ii) as the connections between the conventional expression and its conventional meaning are strengthened, the "less able we are to access alternative, potentially transparent, meanings that are incompatible with the stipulated meaning. In other words, incompatible alternative meanings becomes less transparent" (p.91), while the stipulated or associated meaning becomes more transparent.

To test this hypothesis, the authors had adult students rate very unfamiliar idioms (e.g.: *the goose hangs high* – 'things are going well') on the degree to which the idiom and two associated meanings<sup>a</sup> made sense (roughly, the degree to which the idiom's meaning was analyzable, or at least transparent). Next, in a training phase, participants learned either the idiom's actual figurative sense, or the opposite of the figurative sense (for example, half

<sup>a</sup> For each unfamiliar idiom, the two associated meanings were: a) the figurative paraphrase, and, b) an opposite meaning.

of the participants learned that *the goose hangs high* means 'things are going well', half that it means 'things are going poorly'). After the learning phase, the idioms were again rated for analyzability.

Regardless of whether participants learned the idiom's actual figurative meaning or the opposite meaning, after the learning phase participants rated the learned meaning as being more transparent. That is, familiarity created the intuition of transparency. This would not occur if the meanings of the words comprising the idiom contributed independently to the figurative meaning, as indicated by the semantic decomposition hypothesis. For example, if *blow your stack* is metaphorically transparent and analyzable because of its association with the conceptual metaphor ANGER IS HEATED FLUID IN A CONTAINER, as argued by Gibbs (Gibbs, Bogdanovich, Sykes, & Barr, 1997; Gibbs, 1994; Gibbs & O'Brien, 1990; Nayak & Gibbs, 1990), then the idiom should not be rated as transparent if participants learned that it meant something like 'calmly puff cigar smoke out your nose'. This is because the words comprising the idiom do not relate to the underlying conceptual metaphor. Yet, participants did find opposite meanings as transparent as the presumed metaphorical relationship, if that is what they had been trained on.

Keysar and Bly's experimental design is open to several criticisms: participants may commit to one meaning, and thus later cannot see another meaning as transparent; the learning procedure may have skewed rating judgments, and so forth. However, two points are worth keeping in mind: 1) intuitive judgments of an idiom's transparency and familiarity may be confounded, and this needs to be accounted for in any experimental design; and 2) the use of unfamiliar idioms may therefore provide a 'cleaner' test of the impact of analyzability on idiom comprehension, given Keysar and Bly's findings.

### Summary

The review of the literature to this point has been adult-based. The theories of idiom processing and representation, and the impact of several idiom-internal variables (analyzability and familiarity) derive from, and are tested on, the mental states and the behavior of linguistically mature, figuratively

competent communicators. Almost all of the empirical data derive from studies using only verb phrase idioms, potentially limiting our ability to generalize the outcomes of these studies. Several areas in which the data are in conflict in the adult studies – the role of analyzability on the comprehension of idioms, and of the impact of familiarity on performance and processing, have all also been examined to some degree in developmental studies. Again, the data on each of these factors does not speak with one voice, as noted in the Introduction. Finally, no developmental data are available for generalizing from studies using only verb phrase idioms, to the total, more diverse body of idiom forms. With these points in mind, let us now turn to a review of the developmental evidence.

### Studies of Typical Development

#### Breaking New Ground: Initial Studies in the Development of Idiomatic Competence

As outlined in the Framework, studies of idiom comprehension with adults focus on answering questions about how idioms are processed, where they are represented mentally, and what are their structural characteristics. The earliest investigations of children's understanding of common idioms were primarily concerned with answering the question 'when do children understand the meanings of common idioms?'. A main goal of these early studies was to develop a body of data that would identify when idiom comprehension begins, and when competence reaches adult-like levels. These studies probed the development of comprehension through picture selection (Lodge & Leach, 1975), or a combination of picture selection and verbal explanations of idiomatic meanings (Prinz, 1983; Strand & Fraser, 1979). In these studies, the idiomatic stimuli are verb phrase idioms, each idiom with both plausible literal and figurative glosses (e.g.: *He had a green thumb*; *He faced the music*), and each presented without verbal contextual information. Context, as such, was limited to the visual information required to depict the figurative meaning of each target idiom. The outcomes of these investigations are summarized in Figure 1.

The magnitude of the developmental differences shown by the studies in

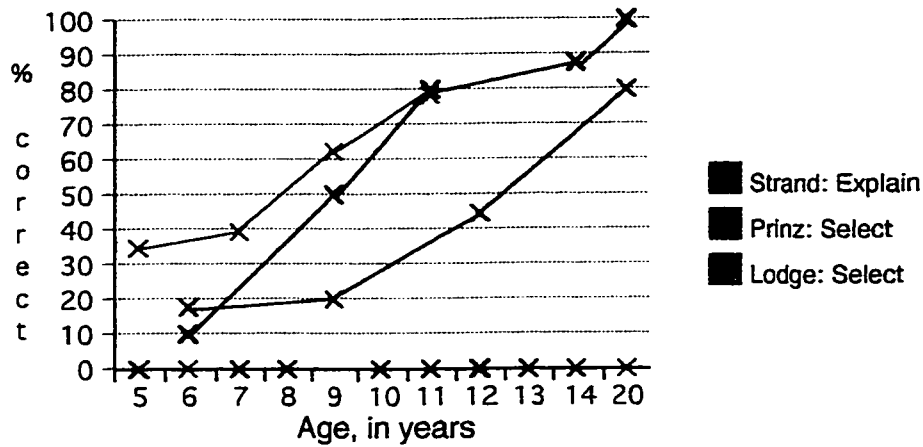


Figure 1: Developmental trends, early studies.

Figure 1 is considerable, particularly in early development. If, for example, 50% correct idiomatic choices' is selected as a reasonable criteria for saying that children are really starting to 'get' idiomatic meanings, then there is nearly a five year difference between Strand and Fraser's (1979) -- 50% correct at 8 years of age, and Lodge and Leach's (1975) -- 50% correct at 13 years old outcomes. Such a finding is surprising, even dramatic. If we posit, based on the methodological information supplied with each study, that:

- 1) the participants are normally developing native English speakers; and,
- 2) the idiomatic stimuli are roughly equivalent in idiom form; and,
- 3) the idioms are equally familiar<sup>9</sup>;

then it is reasonable to assume that other methodological factors influenced the outcomes. Indeed, a variety of task and stimulus factors have been identified that strongly influence when children demonstrate given levels of figurative competence. Thus, a second focus for developmental investigations of idiom comprehension can be characterized as answering the question(s) 'what task and stimulus factors influence idiom comprehension in children?'. Given their potential impact on any developmental study of figurative comprehension, the

<sup>9</sup> The idioms in each study were selected on an intuitive basis as being 'familiar'. The convergent requirement for the stimuli was that the idioms had both plausible figurative and literal interpretations, and these interpretations could be rendered as pictures without captions.

factors that have been identified to date are discussed in the following section.

Figuring out Figurative competence: the influence of task and stimulus factors

Studying the development of idiom comprehension in children presents a three-fold challenge. The first challenge arises from the nature of developmental research. That is, the system itself is in transition, as children rapidly add to, differentiate, and reorganize their linguistic, metalinguistic, and cognitive skills. Differences arising from typical maturation— differences in capacity (of working memory, attending to a task) and level of achievement (ability to process more complex syntax, to formulate and answer direct and inferential questions related to novel narratives) — create challenges in differentiating idiomatic competence from general developmental competence.

A second challenge arises from the uneven patterns observed in typical development. While maturation in each potentially relevant cognitive-linguistic area relates in general to overall development, there is at least local modularity of cognitive sub-systems. That is, development in any given area does not precisely predict development in other areas, and even the modest developmental dissociations seen in typical development require attention and control.

Finally, school children are exposed to idioms through both spoken (Hoffman & Honeck, 1980; Kerbel & Grunwell, 1997; Lazar, Warr-Leeper, Nicholson & Johnson, 1989) and written language (Hollingsed, 1950; Nippold, 1991). While there is no practical means for equating exposure to idioms in conversation, some general means for equating exposure to literature is required, given that children of the same age vary widely in reading skill. Thus, a researcher is faced with a need to specify 'where' the child-as-participant's development 'is' in general; then again more specifically in each area of development; and finally in exposure to literature and academics, since differences in each area have potential impact on the development of idiomatic competence.

These challenges, which risk confounding specific issues in idiom comprehension with more general developmental issues, also offer potential rewards. For example, the ability to observe how children at the very outset of

figurative competence go about solving the problems presented by what are initially novel figurative utterances, and to observe how these solutions might change over time, could provide insights into the general nature of the processing and mental representation of idioms. However, before researchers can effectively take advantage of this proposed developmental dynamic, it is necessary to identify and control the task and stimulus factors that impact the developmental data. These factors are reviewed below first, then the developmental research on analyzability, generalizability and familiarity is examined.

An overview of the developmental research on idiom comprehension allows for two noncontroversial observations related to task and stimulus factors:

1) Children demonstrate figurative competence at an earlier age when the response task requires item selection (picture-selection, item-selection from a verbal or written list) as opposed to verbal explanation (Ackerman, 1982; Ezell & Goldstein, 1991; Gibbs, 1987, 1991; Levorato & Cacciari, 1995; Prinz, 1983); and,

2) Context facilitates figurative comprehension beginning at around six years of age, specifically when the context supports a figurative interpretation of the idiom (Ackerman, 1982; Ezell & Goldstein, 1991; Gibbs, 1987, 1991; Johnson, 1985; Levorato & Cacciari, 1992, 1995; Nippold & Martin, 1989). However, there may not be a facilitating effect for context below six years of age (Abkarian, et al., 1992<sup>9</sup>).

#### Response task demands.

The outcomes of investigations that use both forced-choice and idiom explanations (Gibbs, 1991; Levorato & Cacciari, 1995; Prinz, 1983) are summarized in Figure 2. These results indicate that figurative competence is demonstrated earlier when responses are selected (picture-selection, item-selection from a verbal or written list), then when responses are explained. Indeed, some developmental advantage for item selection is found until the teen years. Given the amount of metalinguistic skill and effort required for

<sup>9</sup> Abkarian et al, 1992; Gibbs, 1987, 1991, and Nippold and Martin, 1989 were the only studies in which direct comparisons between context and no context conditions were run. The other studies compared literal and figurative contexts.

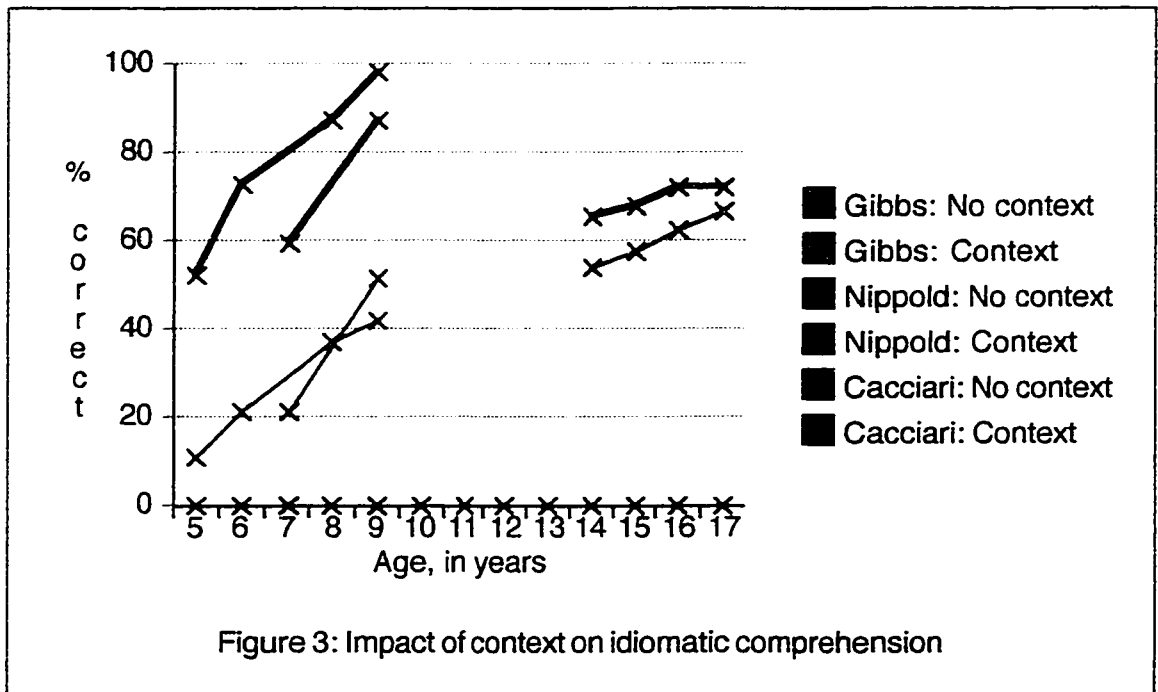
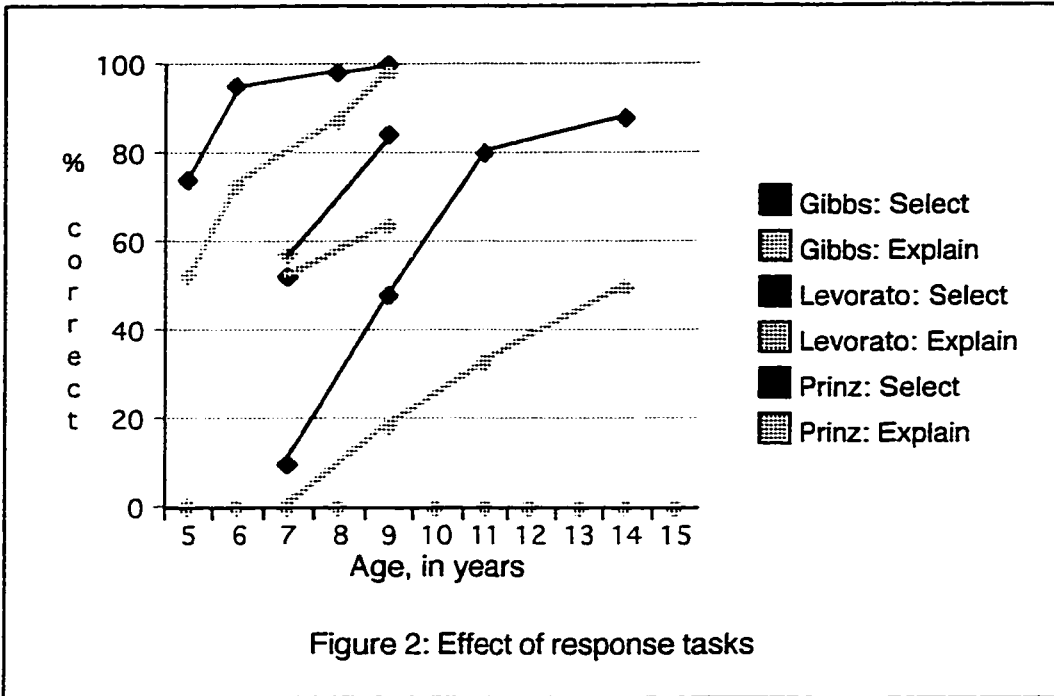
explaining figurative meanings versus indicating meaning through item selection, this finding is not surprising. Item-selection tasks may, therefore, be of greatest use when investigating early developmental trends, since these tasks more sensitively separate figurative knowledge from a child's ability to paraphrase that knowledge coherently and meaningfully. Explanation tasks, on the other hand, might be employed for middle-school and older children, whose metalinguistic competence would support the more challenging explanation tasks.

#### Context Effects.

Two specific context effects have been identified: (1) with-context ('context') versus without-context ('no-context') presentation of idiomatic stimuli; and (2) literal-bias versus figurative-bias narratives in the context condition. The data from the literal- and figurative-bias narratives are not immediately relevant to the issues explored in this dissertation, and will be discussed only as needed to address context issues.

Four different studies evaluated the effects of presenting idioms to children with and without supporting verbal contextual information (Abkarian et al, 1992; Cacciari & Levorato, 1989, 1992; Gibbs, 1991; Nippold & Martin, 1989). The outcomes of three of these investigations are summarized in Figure 3 to illustrate the general trend. As can be seen from Figure 3, in each study children made more idiomatic explanations in the context than in the no-context conditions. The facilitative effect of context appears most powerful in the five-to-nine year age range, but a consistent (albeit modest) effect continues to be seen in the fourteen-to-seventeen year old range.

In discussing the positive effects of context, Levorato and Cacciari (1992,1995) reason that context facilitates children's identification of semantic information which then allows them to develop a coherent interpretation of the idiom within the overall narrative. That is, in young children context is facilitative because it provides information the children need to suspend a literal (word-by-word or phrase-by-phrase) interpretation of an idiom, and to identify the intended figurative meaning that makes sense within the larger text.



However, the facilitative effect of context is not apparent in very young children. Abkarian et al. (1992) found no significant main effect for context when using a picture response<sup>11</sup> measure with 3 - 6 year-old children. Instead, Abkarian found that children made more literal choices with increasing age between three and six years of age. These data suggest that children across this age range select an increasingly 'literalizing' strategy for dealing with contextually ambiguous information.

In sum, the research evidence supports the general conclusion that an appropriate context supports children's ability to determine the figurative meaning of familiar idioms. A context that specifically supports a figurative interpretation is, not surprisingly, more facilitating of a figurative interpretation than contexts that support a literal interpretation. Yet, while context has a powerful influence on comprehension, context does not completely determine children's responses to idioms. For example, older children (10 years old and above) report some figurative meanings even in literal contexts, while very young children provide more literal responses across the age range of 3 - 6 years.

### Stimulus factors in the development of figurative competence

More recently, the focus of developmental research has broadened to include, as is the case with adult studies, the stimulus factors semantic analyzability and idiom familiarity. This represents a shift in focus – from asking when children understand idioms, and what task and context variables influence comprehension – to asking questions about how idioms are processed, based on the impact of semantic analyzability and familiarity on children's comprehension. These issues are reviewed next.

#### Semantic analyzability.

Gibbs (1991) also investigated whether the semantic analyzability of idioms influences the developmental acquisition of idiomatic meanings. Again, Gibbs focused on the dimensions of semantic analyzability identified in his work with adults (Gibbs & Nayak, 1989; Gibbs, Nayak, Bolton, & Keppel, 1989). To

<sup>11</sup> As noted previously, Abkarian et al. placed the depictions of both the literal and figurative meanings in the same response field. As Strand and Fraser (1979) argue, this may induce a literal response bias.

briefly review, these are:

1) Normally decomposable idioms are those in which the "components directly contribute to their overall figurative interpretations" (Gibbs, 1991, p. 613). Examples include *wear out your welcome*, *put down your foot*, and *turn back the clock*;

2) Abnormally decomposable idioms are idioms in which "each part does not by itself refer to some component of the idiomatic referent, but only to some metaphorical relation between the individual part and the referent" (p. 614). Examples include *sitting on pins and needles*, *breathe down your neck*, and *roll out the red carpet*;

3) Unanalyzable or nondecomposable idioms are idioms for which the "individual parts do not contribute individually to (the) figurative meaning" (p. 614). Examples include *turn over a new leaf*, *beat around the bush*, and *throw in the towel*.

Gibbs predicted that analyzable idioms would be easier for children to learn than unanalyzable idioms. This would occur because: (a) the components of analyzable idioms do systematically contribute toward the overall meaning; and (b) children would be likely to access the semantic representations of each component, then combine these meanings according to the routine compositional rules of the language. On the other hand, using a compositional strategy would hinder comprehension of unanalyzable idioms. This would be so because children could not use the meanings of the individual words to decompose the unanalyzable idioms into their underlying figurative meanings, and there would only be general contextual cues to their meanings.

To investigate this hypothesis, Gibbs asked 5, 6, 8 and 9 year-old children to interpret verb phrase idioms in both with-context and no-context conditions (see Figure 3 for the between-contexts outcomes). After each stimulus item was read, the children were first asked to explain the meaning of the final sentence which contained the target idiom. Next, participants were read two possible explanations of the target sentence and then selected the one that best represented the meaning of the target.

Results for the children's explanations showed significant effects for age; for analyzability (more correct explanations for analyzable idioms); and for

context, with more correct idiomatic explanation in the with–context condition. In the with–context condition both kindergartners and first graders gave significantly more correct explanations for normally analyzable idioms than for either abnormally or unanalyzable idioms. In the no–context condition participants at every grade level provided significantly more correct explanations for normally analyzable than either abnormally or unanalyzable idioms.

Results from the forced choice questions showed similar effects for age, context, and analyzability. High levels of accuracy were observed in selection of the figurative paraphrase in both context conditions and for both analyzable and unanalyzable idioms for all grades beyond kindergarten (90%+), with ceiling effects apparent by Grade 3. Due to this ceiling effect, only the kindergartners showed a significant main effect for analyzability, selecting significantly more analyzable than unanalyzable idioms. In the no-context condition children at each grade chose significantly more correct idiom paraphrases for the normally analyzable idioms than for either the abnormally or unanalyzable idioms. In the no–context condition the kindergartners and first graders were more likely to select the literal paraphrase for each idiom type, while even fourth graders had not convincingly abandoned the literal paraphrase for the unanalyzable (39%) idioms.

Gibbs (1991) argues that these results indicate that children use a compositional analysis when encountering idiomatic expressions. Children use their "intuitions as to how the parts of idioms contribute to their overall figurative meanings" ( p. 619) to aid them in learning the figurative meanings of analyzable idioms. This is a strong claim regarding semantic analyzability, and if consistently replicated can add significantly to our understanding of children's mastery of particular idioms at different developmental stages. However, Gibbs' 1991 outcomes are potentially limited by:

a) the researchers attempting "to convey the idiomatic meaning of each phrase when it was read" (p.615), but not defining how the 'conveying' was done. While idiomatic meanings are apparently easier to understand when spoken using richer prosody (Van Lancker & Canter, 1981), the deliberate use of heightened prosody seems a difficult to operationalize variable; and

b) the use of unbalanced numbers of analyzable (9 normally analyzable + 6 abnormally analyzable) and unanalyzable (5) idioms in the two context conditions. With so few unanalyzable idioms in each context/no-context list it may be that the results are due to the children's being primed (by the sheer numbers of analyzable idioms) to use a word-by-word analysis.

Two studies by Nippold (Nippold & Rudzinski, 1993; Nippold & Taylor, 1995) also examined the impact of transparency/analyzability, and of familiarity on older students comprehension of idioms. The idioms, the procedures for identifying the idiom's transparency and familiarity, and the associated story contexts are common to both studies. The procedures and outcomes relevant to transparency are presented in this section, the material on familiarity in the following section.

First, the transparency of 100 familiar and unfamiliar verb phrase idioms was determined by asking both 12th graders and university students to rate<sup>12</sup> how closely the literal and non literal meanings of each idiom were related. While there was a high correlation between the high school and university students' ratings ( $r = .86$ ), the adolescents rated the idioms as significantly more transparent than the adults did. Nippold reasoned that this might be because with greater experience the adults' understanding of the idiomatic meanings was more fixed and conventional, and hence less transparent (but see Keysar & Bly, 1995, for a different argument on transparency-familiarity). On the other hand the less experienced high school students might still be in a period when they are more actively analyzing each idiom's components for individual meanings, and this might cause them to rate the idioms as more transparent.

In Nippold's initial investigation (Nippold & Rudzinski, 1993), students in grades 5, 8 and 11<sup>13</sup> explained in writing the meanings of high-, moderate- and low-familiarity, transparent and opaque verb phrase idioms, presented within idiom-bias contexts. Significant, moderate correlations between transparency and the accuracy of the explanation questions were obtained for grades 8 ( $r = -.47$ ) and 11 ( $r = -.53$ ), and for all grades combined ( $r = -.49$ ), but not for grade 5 ( $r = -.35$ ), with more accurate figurative explanations provided for transparent

<sup>12</sup> Ratings were made on a 3-point scale (1 = literal and non literal meanings are closely related; 3 = literal and non literal meanings are not related).

<sup>13</sup> N = 50 students in each grade. All experimental procedures for both studies were administered in classroom groups within the classroom setting.

than for opaque idioms. Given Gibbs (1991) findings that transparent idioms are easier for students in kindergarten through grade 4 to understand than opaque idioms in an explanation task, it is not clear why, in Nippold's study, there was not a significant effect for analyzability at grade 5 (which could reasonably be inferred from Gibbs' outcomes), but there was for the grade 8 and 11 students.

In a subsequent investigation, Nippold and Taylor (1995) again investigated the impact of transparency on the comprehension of idioms, but used item-selection responses in place of the explanation tasks in order to also compare the impact of response mode. Again, transparency facilitated comprehension: There were significant, moderate correlation coefficients between comprehension and transparency for each grade (grade 5 :  $r = -.48$ ; grade 8:  $r = -.47$ ; grade 11:  $r = -.54$ ), and for all grades combined ( $r = -.52$ ), with more correct figurative choices for transparent idioms than opaque idioms.

In contrast with the results described above, Abkarian (Abkarian, et al., 1992) found no significant effect for transparency on idiom comprehension. Abkarian used a picture selection response task to study idiom comprehension in 3 - 6 year old children. The children were tested in two context conditions, context and no-context, with order of presentation balanced. No significant correlation was found between the children's responses and adults' later ratings of the transparency of each idiom ( $p = .15$ , using a Spearman rank order correlation procedure), indicating that there was no impact for transparency.

Finally, Abrahamsen and Burke-Williams (1996) investigated idiom comprehension in both typically developing and language delayed third- and fifth-grade students. Comprehension of both transparent and opaque idioms was assessed for each group. No significant differences in comprehension of transparent versus opaque idioms was found in either the typically developing or language delayed groups. While the language delayed group was less successful overall in comprehending the figurative meanings of idioms at each grade level, the transparency of the idioms per se did not aid either group's successful interpretation of idiom meanings.

While some methodological criticisms can be made of the Abkarian study (idiom transparency was rank-ordered, not scaled), for our purposes the central issue is, for both the Abkarian et al. and Abrahamsen and Burke-Williams

studies, that no significant effect for transparency was found for these familiar idioms. This lack of an effect predicted by the semantic decomposition hypothesis, and supported by several other studies (see Table 2 for a summary of the developmental research), lead us to conclude that additional investigations into the impact of analyzability on children's comprehension of idioms are required in order to disambiguate these conflicting results.

Table 2

Effects of Analyzability on Comprehension in Typically Developing Children, Reported in 8 Studies.

Study	Age/Grade (G)							
	3+4	5	6	8	9	G5	G8	G11
Gibbs (1991) <sup>a</sup>		1/1*	1/1	0/1	0/1			
Gibbs (1991) <sup>b</sup>		1/1	0#/1	0#/1	0#/1			
Nippold & Rudzinski (1993) <sup>a</sup>						0	1	1
Nippold & Taylor (1995) <sup>b</sup>						1	1	1
Abrahamsen & Burke-Williams (1996)				0		0		
Abkarian, et al (1992)	0	0	0					

0 = no significant effect for analyzability on comprehension

1 = significant effect for analyzability, with comprehension of analyzable idioms > comprehension of unanalyzable idioms at the specified age.

\* = **with-context/no-context**

# lack of significant effect for analyzability may be due to ceiling effect in the forced-choice task

<sup>a</sup> Explanation task

<sup>b</sup> Forced-choice response task

In recent work, Gibbs (Gibbs, 1992, 1994; Gibbs, Bogdanovich, Sykes, & Barr, 1997; Gibbs & Nayak, 1991; Gibbs & O'Brien, 1990) has argued strongly for the metaphoric mediation between the surface reading of the idiom and its underlying figurative sense, moving away from dividing among 'normally,

abnormally, and non-decomposable idioms'. This tripartite division is not maintained elsewhere in the literature, and for the purposes of this dissertation, a more functionalist view of analyzability is used, a view more in keeping with most other investigations into semantic analyzability. Here, we will conceptualize idioms as extending along a continuum of analyzability, from transparent/analyzable to opaque/unanalyzable idioms, without positing a third category of 'abnormally analyzable' idioms. Using idioms equated for familiarity, we propose that analyzable/transparent idioms are those for which adults reliably report a direct relationship between the words comprising an idiom, and the idiom's figurative meaning as presented in widely used dictionaries of idioms. Examples of analyzable idioms, as operationally defined here (and rated using the procedures to be outlined in 'Methods', below), include *the fat of the land*, *asleep at the switch*, and *ask for the moon*. Unanalyzable/opaque idioms are those which adults reliably report no obvious relationship between the words comprising the idioms, and the idiom's figurative meaning as derived from widely used dictionaries of idioms. Examples of unanalyzable idioms, again as operationally defined here, include *a pig in a poke*, *three sheets to the wind*, and *carry coals to Newcastle*.

#### The effects of familiarity on idiom comprehension in children.

The preponderance of the evidence from idiom studies with adults, as reviewed in Theories of Idiom Processing and Representation, above, indicates that familiar idioms show processing speed and response accuracy advantages over unfamiliar idioms. However, as several studies have shown, it is not entirely clear just 'what' is familiar in familiar idioms (Forrester, 1995), and whether transparency ratings are in some sense confounded by familiarity (Keysar & Bly, 1995). In the developmental literature much, but not all, of the evidence likewise supports the conclusion that children are more successful in identifying the figurative meanings of familiar idioms than of unfamiliar idioms. However the developmental evidence does indicate that comprehension may be more directly influenced by context than by familiarity, at least for younger children. Since considerations of context typically dovetail with arguments supporting analyzability, while familiarity/experience evidence can be used to support viewing idioms as lexicalized, it is to these findings that we now turn.

Nippold argues for a language experience hypothesis regarding idiomatic competence, in which the amount of exposure to figurative forms is critical to the development of idiomatic competence. If familiar idioms were identified more accurately than unfamiliar idioms in probes of idiom comprehension, then this finding would be seen as supporting the language experience hypothesis, since amount of exposure essentially creates the rating of 'familiar'.

To test their reasoning, Nippold and Rudzinski (1993) asked the adolescents and adults who participated in the transparency ratings (above) to rate idioms for familiarity. From these ratings, three groups of idioms: high<sup>14</sup>, moderate and low familiarity ( $n = 8$  per group) were extracted and used as stimulus items with accompanying story contexts. The idiomatic stimuli were then presented as outlined above under Analyzability to students in grades 5, 8, and 11 (again, who did not participate in the ratings), who explained the meaning of each terminal sentence in writing.

Significant results were found for: (i) grade, with accuracy increasing from grade 5 to grade 11; and (ii) familiarity, with the most correct figurative answers for familiar idioms, the least for the low-familiarity idiom. Grade x Familiarity interactions were also obtained. Post hoc testing indicated that the differences between the high- and low-familiarity idioms were significant at each grade level. For example, 5th grade students provided correct explanations for 31% of low-familiarity idioms, vs 66% correct explanations for high-familiarity idioms. Eleventh grade students provided correct explanations for only 54% of low-familiarity idioms, vs 87.% of high-familiarity idioms.

Next, Nippold combined the results from all students, to derive a difficulty ranking for the 24 target idioms (for example, 93% of all students answered *strike the right note*, and *keep up one's end* correctly; only 7% answered *lead with one's chin* correctly). The 'difficulty rankings' were then correlated with the familiarity scores. Significant moderate correlations were found for each grade level ( $r_s$  -.51 to -.54), indicating that familiarity is a moderately good predictor of performance on the idiom comprehension task. In a follow-up study using a forced-choice (item-selection) response task, the same basic pattern of results

<sup>14</sup>'High' average rating = 1.33; low-familiarity rating = 4.29, where 1 = very familiar, 5 = never heard or read it.

was reported, with higher accuracy for the item-selection task at each grade than for the explanation task (Nippold & Taylor, 1995).

Nippold argues that these outcomes all support the language experience view of idiom comprehension, in which the amount of exposure is uniquely predictive of accuracy in idiom comprehension tasks.

In contrast to Nippold's findings, Levorato and Cacciari (1992) did not find that familiarity significantly facilitated children's figurative comprehension of transparent idioms. Levorato and Cacciari presented 7- and 9-year-old children with familiar and unfamiliar idioms in both context and no-context conditions. The familiarity of each idiom was first determined by asking 152 elementary school teachers to rate, on a five - point scale, how frequently they thought children were exposed to each idiom. Given evidence from other studies indicating the high frequency with which children are exposed to idioms within school (Hollingsed, 1950; Kerbel & Grunwell, 1997; Lazar, et al., 1989; Nippold, 1991), this methodology presents with reasonable face validity.

In the context condition, the children were presented with idioms in both literal and figurative-bias short stories. Following each short story, comprehension was assessed by having each child answer a question about the character's activity in the story by selecting one of three alternatives in a multiple-choice format. The three choices always contained: (i) the figurative idiomatic interpretation; (ii) a paraphrase of the literal meaning; and (iii) an alternate but plausible response unrelated to either the figurative or literal sense of the idiom.

No main effect for familiarity was found (as represented in Figure 3, a main effect for context was obtained) on the acquisition of figurative competence. In the context condition 7-year-old children selected the idiomatic choice for familiar idioms in 50% of trials in the idiomatic-bias condition, and for unfamiliar idioms in 34% of trials. The 9-year-old children made the same selection in 64% of trials with familiar idioms, 68% for unfamiliar idioms. A similar pattern of results was obtained in the no-context condition, but with lower overall accuracy. The 7-year-old children selected the idiomatic choice in 26% of trials with familiar idioms, 12% for unfamiliar idioms. The 9-year-old children selected the idiomatic choice in 48% of trials for familiar idioms, 48% for

unfamiliar idioms.

Levorato and Cacciari argued that these results supported the hypothesis that familiarity per se is not as potent a variable as context in supporting children's comprehension of figurative meanings. They argue that context provides information that is essential for processing small elements of language within the broader purpose of making overall sense of larger units of discourse or text. In this view, exposure (per the language experience hypothesis) is not as important as the processing strategies children use. Children select processing strategies which help them to "build a coherent semantic representation of the story and to integrate the local information given by the words composing the idiom into the global meaning conveyed by the narrative" (Levorato & Cacciari, 1992, p. 431). In their view, exposure may still have a role in acquiring figurative competence, but it is a role that is secondary to the child's efforts in making sense of the overall text or conversation. Processing strategies, primarily working from narrative meaning and coherence to the more local meaning of the idiom, are how children come to understand the figurative meanings of idioms, and of unfamiliar language in general.

A conservative summary of this portion of the literature suggests that *on the face of it*, research into the development of idiom comprehension needs to control for the familiarity of the idiomatic stimuli. However, as the research summarized in Table 3 outlines, the impact of familiarity is apparently not uniform across all ages or in particular stimulus conditions. Younger children seem to depend more heavily on context than on familiarity per se (Levorato & Cacciari, 1995), while older children (and adults – see Kemper, 1986; McGlone, et al, 1994) use their prior experience with familiar idioms to solve the question of 'what does this all mean, anyway?' (Nippold & Rudzinski, 1993; Nippold & Taylor, 1995). Finally, some evidence from adult studies cautions us to consider the exact nature of the familiarity that the researcher will want to control for. Forrester (1995) reports that 'what' is familiar in familiar idioms is their meaning, not their form (Forrester, 1995), and Keysar and Bly (1995) that the perception of analyzability may be more a function of familiarity than of the semantic analyzability of the parts, per se.

Table 3

Effects of Familiarity on Comprehension in Typically Developing Children and Adults. Reported in 8 Studies.

Study	Age/Grade					
	7	9	G5*	G8	G11	Adult
Nippold & Rudzinski (1993)			1,2	1,2	1, 2	
Nippold & Taylor (1995)			1	1	1	
Levorato & Cacciari (1992)	0	0				
Forrester (1995)						3, 0 <sup>1</sup>
Kemper (1986)						3
McGlone, et al (1994)						3
Schraw, et al (1988)						0 <sup>2</sup>
Keysar & Bly (1995)						3, 0 <sup>3</sup>

0 = no significant effect for familiarity on comprehension

0<sup>1</sup> = reading time advantages found for both familiar idioms and the 'semantically equivalent' paraphrases of familiar idioms, raising the question 'what's familiar?'

0<sup>2</sup> = familiarity does not affect reaction time; does affect preference for selection of a figurative over a literal interpretation

0<sup>3</sup> = familiarity and transparency interact, with familiarity generating an 'illusion' of transparency

1 = significant effect for familiarity, with comprehension of familiar idioms > comprehension of unfamiliar idioms at the specified age.

2 = significant Grade x Familiarity interactions

3 = significant reaction time advantage for familiar over unfamiliar idioms

$p \leq .05$  in all studies reported above

Generalization: Can Verb Phrase Idioms Represent *The Whole Shooting Match*?

In the section 'Theories of Idiom Processing and Representation' it was noted that almost all of the empirical data in adult studies comes from studies using only verb phrase idioms. A review of available developmental studies

indicates that, for those studies that either provide or specify the idioms used as stimuli, 95-99% of stimuli are verb or verb phrase idioms (see, for example: Abkarian, et al.<sup>15</sup>, 1992; Ackerman, 1982; Ezell & Goldstein, 1991; Gibbs, 1991; Lodge & Leach, 1975; Nippold & Rudzinski, 1993; Nippold, Taylor, & Baker, 1996; Prinz, 1983; Strand & Fraser<sup>16</sup>, 1979 and others). As the review of both the adult and the developmental literature has demonstrated, there certainly are clear advantages to identifying, isolating and holding in common among studies various crucial variables. In this sense, the almost exclusive use of verb phrase idioms has arguably been beneficial to research on idioms, permitting some kind/s of comparisons among studies which otherwise show considerable diversity.

At the same time, the nearly exclusive use of verb phrase idioms potentially limits our ability to generalize the outcomes of both the adult and developmental studies to other forms. An overview of common dictionaries of American idioms indicates that only about 50% of idioms are verb or verb phrase idioms. Further, as Nippold observes, idioms present in a variety of forms and serve a variety of functions. As examples,

"Syntactically, idioms can range from single words (e.g., *nosey*) to multiword productions that include noun phrases (e.g., *chip on his shoulder*), verb phrases, (e.g., *lost his heart*), adjective phrases (e.g., *green with envy*), adverb phrases (e.g., *like clockwork*), and independent clauses (e.g., *the cat's got his tongue*)....semantically, idioms can (be categorized) as foods (e.g., *sour grapes*), animals (*dark horse*), clothing (e.g., *keep one's shirt on*)....(and can serve a variety of pragmatic functions, such as) to admonish (e.g., *Stop beating around the bush*), to praise (e.g., *The detective hit the nail on the head*), and to express sarcasm (e.g., *The princess has a hard row to hoe*) and humor (e.g., *Clean socks were scarcer than hens' teeth*)." (Nippold, 1991, p. 100).

Nippold has additional, and entertaining examples of idiomatic diversity in her 1991 article. Beyond even the diversity of idioms identified by Nippold are 'quirky' anomalous idioms, such as *which is which* and *by and large*, for which no developmental parameters are available. Central to this discussion is the point that idioms come in many 'packages' of form and content, and are put to

<sup>15</sup> Nine of ten idioms.

<sup>16</sup> Seventeen of twenty idioms.

diverse uses. Given this, developing theories of the acquisition of figurative competence based on only one form, the verb phrase idiom, indeed seems a leap of faith.

Given the syntactic, semantic and pragmatic diversity demonstrated by idioms, it would seem reasonable at this stage of developmental research to add breadth to the empirical data base, to provide a broader foundation from which hypotheses regarding developmental and comprehension processes can be derived. This dissertation proposes to investigate one area of diversity, the syntactic form of idioms per se. It is to this proposal, and to the questions motivating this dissertation, to which we now turn.

### Summary

As first noted in the 'Introduction', the objectives of the research presented in this dissertation are three-fold. The first objective is to reexamine the unresolved issue of analyzability, by using only unfamiliar idioms when probing the development of figurative comprehension. To present the issue of semantic analyzability, we first reviewed the recent history of research regarding idiom comprehension, then proposed a 'working definition' for idioms to be used throughout this paper. Next, current issues in idiom research were outlined and identified as evolving along three primary axes -- issues regarding how idioms are processed, mentally represented, and how idioms behave internally, with detailed reference to their semantic flexibility and analyzability. Three theories of idiom processing in adults -- literal-first, simultaneous, and figurative-first, were discussed, and it was concluded that no single theory adequately explains and predicts how adults process idioms.

One possible reason for why no single theory of idiom processing and/or mental representation demonstrates explanatory adequacy may be because our understanding of the internal behavior of idioms has also evolved only gradually. Current hypotheses regarding the internal dynamics of idioms now include both issues of semantic analyzability and familiarity. One theory of the internal behavior of idioms proposes that idioms are semantically flexible, and can be analyzed into meaningful parts. This hypothesis of analyzability, the

Semantic decomposition hypothesis, developed by Gibbs and associates (Gibbs & Nayak, 1989; Gibbs, Nayak, Bolton & Keppel, 1989; Gibbs, Nayak, & Cutting, 1989), proposes that idioms behave differently because for some idioms, the individual parts of the idioms contribute independently to the figurative meaning. Gibbs hypothesizes that both adults and children can analyze these parts to help derive the figurative meanings of 'analyzable' idioms. For other, 'unanalyzable' idioms, the parts do not contribute to the figurative meanings, and the idiomatic meaning has to be learned directly.

While the semantic analyzability of idioms does account for some of the identified behavior of idioms, the evidence to date does not uniformly support the semantic decomposition hypothesis. In one adult study by Gibbs (Gibbs, Nayak & Cutting, 1989), not all of the data supported the authors' own predictions. In a separate study of cued recall, no predictable effect for analyzability was found (Cutting & Bock, 1997). Applying the notion of semantic analyzability to developmental research, several studies (Gibbs, 1991; Nippold & Rudzinski, 1993; Nippold & Taylor, 1995) reported that students identified the figurative meanings of analyzable idioms more often than unanalyzable idioms. In contrast, studies by Abkarian (Abkarian, et al., 1992) and Abrahamsen (Abrahamsen & Burke-Williams, 1996) reported no significant differences between analyzable and unanalyzable idioms in children's acquisition of idiomatic meanings.

Given these conflicting outcomes, one goal for this dissertation is to add into the balance of the debate on analyzability, data for which a principal confound -- familiarity -- has been effectively controlled. The rationale for controlling for familiarity by using only unfamiliar idioms is based on this argument: By using (unfamiliar) idioms from which much of the experiential 'baggage' has been stripped, a basic view of the impact of analyzability can be obtained. This would be so because, with (very) unfamiliar idioms, students will have only contextual information from the accompanying stories to use in glossing the idiomatic meanings of all the target idioms. However, if analyzability does aid in comprehending figurative meanings, then students should be more successful in identifying the figurative meanings of analyzable than unanalyzable idioms. This would occur because for analyzable idioms

students can use both contextual cues and the semantic information from the components of each analyzable idiom in identifying the idiomatic meaning. On the other hand, for unanalyzable idioms students would have access only to the general contextual information to aid in identifying the figurative meaning. Presumably, when confronted with new and unknown complex phrasal units, two sources of data are better than one for mapping new meanings, and analyzable idioms should be understood more readily than unanalyzable idioms.

The second objective of this dissertation is to investigate whether, and to what degree, the developmental data on idiom comprehension, derived almost exclusively from verb phrase idioms, might generalize to other idiom forms. The issue is framed with reference to the impact of the syntactic form of idioms on children's developing comprehension competence. The goal is to determine whether other commonly occurring idiom forms (noun phrase and adjective phrase idioms) behave in the same manner developmentally as verb + noun phrase (verb phrase) idioms. Again, as Nippold (1991) observes, the syntax of idioms ranges from single words (*nosey*) to independent clauses (*the cat's got his tongue*), and include verb, noun, adjective and adverb phrase idioms, and idioms serve a variety of semantic and pragmatic functions. Despite this diversity, only verb phrase idioms have been used systematically in developmental research to date. Thus, it is necessary to establish the degree to which data on verb phrase idioms generalizes to other idiom forms.

The third objective of this dissertation will be to determine whether semantically analyzable and unanalyzable idioms behave differently based on their idiom form. That is, is there a main effect for analyzability or for form, or are there interaction effects among analyzability and form? As a corollary to the exploration of this topic area, the behavior of syntactically and semantically anomalous idioms – idioms such as *which is which* and *come a cropper*, will also be described. This will be done in order to develop some initial developmental data regarding how these fundamentally 'rule-breaking' idioms behave within the experimental environment described in the following chapter, Methods. These 'oddball' idioms are, despite their unusual construction, frequently occurring in at least certain settings (including elementary school;

see Lazar, et al., 1989). One would predict, based on Gibbs' hypotheses, that comprehension of anomalous idioms would be a late-developing achievement, since these idioms are seemingly quite opaque. Because these forms have never been researched developmentally, actual outcome remains an empirical question.

#### Statement of the Problem:

Given the above as background, the questions which motivate this investigation can be summarized as follows:

1. Do children in grades two, five and eight demonstrate differential success in identifying the figurative meanings of analyzable and unanalyzable, unfamiliar verb phrase idioms within contexts that support a figurative interpretation?

Based on the preponderance of the developmental data to date, the prediction is that analyzability will influence idiomatic comprehension, with more children able to identify and/or explain the accurate figurative meaning of analyzable than unanalyzable idioms.

2. If analyzability differentially affects the acquisition of figurative competence at any grade, does the effect of analyzability vary based on grade?

The developmental data indicate that children use a more literalizing strategy around six and seven years of age, with many common idioms understood by nine or ten years of age. Thus, a greater effect for analyzability will be evident in grades two and five. By eighth grade, students will use their general facility with figurative forms and ability to use context to a degree that they can 'fill in the blanks' with even unfamiliar, unanalyzable idioms.

3. Does the syntactic form of an idiom differentially impact the acquisition of figurative competence by children in grades two, five and eight? Specifically, is the developmental course of figurative competence for noun phrase and adjective phrase idioms similar to the developmental course for verb phrase idioms across grades two to eight?

4. Does the analyzability of noun phrase and/or adjective phrase idioms differentially impact the acquisition of the figurative meanings of these idiom forms?

It is hypothesized that analyzable noun phrase idioms can provide more coherent, or at least readily visualized information for analysis (e.g.: *on the horns of a dilemma*; *a bee in her bonnet*) than analyzable adjective phrase idioms (e.g.: *out of the swim*; *bred in the bone*). If this is accurate, then the figurative meanings of analyzable noun phrase idioms will be identified more successfully than adjective phrase idioms.

5. If analyzability differentially affects the acquisition of figurative competence for noun and/or adjective phrase idioms at any grade, does the effect of analyzability vary based on grade?

6. Are there any interaction effects among analyzability, idiom form, and grade?

7. How does children's acquisition of ill-formed or anomalous idioms compare with their acquisition of well-formed idioms?

These idioms are true anomalies, and as such may benefit from their perceptual salience in conversation. It is predicted that, although they are likely unanalyzable overall, they will be comprehended more readily than other unanalyzable forms, due to an 'odd-ball' effect, but that students will not identify these forms with as much accuracy as analyzable idioms in general.

## CHAPTER TWO

### Methods

#### Subjects

Thirty-six students participated in this study: 12 in each of Grade 2, 5 and 8 were included in the experimental tasks and the data analysis. All students attended schools in the Seattle metropolitan area. The age range of Grade 2 - Grade 8 was selected because an overview of research with typically developing students shows strong growth in the ability to identify and explain idiomatic meanings across this period, while minimizing likely floor or ceiling effects on the response tasks (Ackerman, 1982; Cacciari & Levorato, 1989; Nippold & Rudzinski, 1993; Nippold & Taylor, 1995; Nippold, et al., 1996; Prinz, 1983; Strand & Fraser, 1979).

Students were identified for participation through a two-stage screening and selection process (see specific criteria, below). Forty-seven students were referred by their teachers for initial screening based on the likelihood that they would meet the described screening criteria. Forty-five students (20 male, 25 female) met screening criteria and were given the selection measures in a single session at each student's school. A total of 38 students met all selection criteria. One student was subsequently dropped from data analysis when it was discovered ex post facto that he was repeating the grade. A final student (with the lowest combined WISC-R and CELF-R scores) was dropped from the data analysis in order to balance the number of students at each grade. This left a total of 36 students (14 males, 22 females), 12 students at each of Grade Two (G2), Grade Five (G5), and Grade Eight (G8) (see Table 4 for summary demographic data). All students came from families in which at least one parent was regularly employed.

Screening Criteria. The children selected for further assessment were first required to meet the following inclusionary and exclusionary criteria to enhance group by grade homogeneity:

Table 4

Participant Demographic Data, by Grade

Grade	Mean age <sup>1</sup>	range <sup>1</sup>	gender <sup>2</sup>	n <sup>3</sup>
2	8;04	7;09 - 8;10	3:9	12
5	11;04	10;10 - 12;03	7:5	12
8	14;02	13;09 - 14;08	4:8	12

<sup>1</sup> age in years; months

<sup>2</sup> males:females

<sup>3</sup> students per grade

Inclusionary criteria:

- 1) Each student was in his/her first year at each grade level;
- 2) Each student demonstrated achievement at grade level (A to C range for letter grades) in reading and math skills, based on teacher report, most recent report card results, and results of any recent (within one year) standardized achievement tests on file.

To control for the linguistic diversity of the Seattle metropolitan area, a final common criteria was:

- 3) Each student spoke English as the first language of the home.

Exclusionary criteria (adapted, in part, from Stark & Tallal, 1981)

No child demonstrated:

- 1) A known history of significant hearing loss, including present use of hearing aid(s) or current reported occurrence of otitis media (parent and teacher or nurse report);
- 2) A known or evident neurological impairment (teacher report, clinical observation);
- 3) A known, severe behavioral or personality disorder (clinical observation, teacher report);
- 4) A prior history of speech or language therapy (teacher report, student

school file).

Students who passed the screening criteria above were then assessed for compliance with the selection criteria below, using standardized probes of cognitive-linguistic development.

Selection Criteria. The children chosen to participate in the experimental tasks met the following developmental parameters:

- 1) Oral language development within  $\pm 1.0$  standard deviations around the mean (standard scores between 7.0-13.0), as estimated by averaging each student's scores from the Paragraph Comprehension and Sentence Recall subtests from the Clinical Evaluation of Language Fundamentals--Revised (CELF--R), (Semel, Wiig & Secord, 1987);
- 2) Estimated verbal I.Q. within  $\pm 1.0$  standard deviation around the mean (standard scores between 7.0-13.0), as estimated by averaging each student's scores from the Vocabulary and Comprehension subtests from the Wechsler Intelligence Scale for Children--Revised (WISC--R), (Wechsler, D., 1974).

The CELF-R was also used to establish that students could:

- a) remember material in immediate verbal memory (CELF-R, Sentence Recall). Immediate verbal memory was required in the forced choice task, in which students responded to verbally-presented multiple-choice items; and,
- b) remember and comprehend salient information from short stories (CELF--R, Paragraph Comprehension). Retention and comprehension of information from short stories was required in both the item selection and explanation response tasks.

The WISC-R measures were used to ensure that:

- c) students' knowledge base for common nonidiomatic words was adequate to support, in general, comprehension of the directions, short stories, and forced-choice items (WISC--R, Vocabulary); and,

- d) students' ability to explain core ideas associated with routine social concepts (WISC--R, Comprehension) was developing within normal limits.

These estimators of I.Q. are important (a) in establishing group homogeneity, given the relationships between I.Q. and idiom comprehension within the target age range as reported by Ezell and Goldstein (1992); and (b) in helping to

isolate any developmental challenges in identifying and explaining figurative meanings for unfamiliar idioms from more general issues of cognitive/linguistic competence.

To determine whether there were any significant differences in general language or intellectual development among the three grades, one-way Analyses of Variance (ANOVA) were computed for the averaged CELF-R and WISC-R standard scores (summary statistics for the selection measures are presented in Table 5). There were no significant differences between the grades for either language ability,  $F(2, 33) = 1.2765$ ,  $p = .2924$ , or for estimated verbal I.Q.,  $F(2, 33) = .5073$ ,  $p = .6068$ .

Table 5

Students' Clinical Evaluation of Language Fundamentals-R<sup>1</sup> and Wechsler Intelligence Scale for Children-R<sup>2</sup> performance data, by grade

measure	<u>CELF-R</u>		<u>WISC-R</u>		<u>n<sup>3</sup></u>
	Mean	SD	Mean	SD	
Grade					
2	9.83	1.95	10.79	1.23	12
5	9.21	1.78	10.29	1.34	12
8	<u>10.33</u>	<u>1.42</u>	<u>10.38</u>	<u>1.33</u>	12
<u>M</u>	9.79	1.73	10.49	1.28	

<sup>1</sup> Average of Sentence Recall and Paragraph Comprehension subtest standard scores.

<sup>2</sup> Average of Comprehension and Vocabulary subtests standard scores.

<sup>3</sup> students per grade

### Stimuli

A total of 34 idioms were used for the explanation and recognition portions of the comprehension tasks. The stimuli consisted of eight items in

each of four idiom forms: Verb Phrase (VP), Noun Phrase (NP), Adjective Phrase (ADJP) and Anomalous (ANOM) idioms. In addition, two idioms were used for training participants on the experimental tasks. The idiom forms VP, NP and ADJP each included four idioms judged as Semantically Analyzable (SA) and four idioms rated as Semantically Unanalyzable (SU) idioms (see Stage 2 - Analyzability Ratings, for how these ratings were obtained). The eight ANOM idioms did not meet criteria for either familiarity or analyzability, and will be treated separately in the data analysis. Selection of the idiomatic stimuli was done in these three stages:

#### Stage 1 - Familiarity Ratings.

The experimental design called for the use of only unfamiliar idioms as stimulus items. Five undergraduate students were paid to independently judge the familiarity of 176 idioms<sup>17</sup>. The initial pool of 176 idioms was drawn from two primary sources. First, idioms identified from previous experimental work as unfamiliar and including both semantically analyzable (SA) and semantically unanalyzable (SU) idioms (Gibbs, 1991; Gibbs & Nayak 1989) were used. Second, idioms were chosen from the first edition of a commonly used dictionary of American idioms (Boatner & Gates, 1966), and were judged on an ad hoc basis by the experimenter as likely to be unfamiliar and SA-SU. This older edition of A Dictionary of American Idioms was used, given that it would be more likely to contain idioms no longer in current usage, and hence less familiar (see Keysar & Bly, 1995, re: identifying resources for selection of unfamiliar stimuli).

From these sources, lists of idioms were created for the participants to judge. Each idiom list ( $N = 176$ ) consisted of 40 each of VP, NP, ADJP and ANOM idioms. Each idiom form's list was then tentatively balanced for SA and SU idioms based on two speech-language pathologists intuitions on the analyzability (SA - SU) of each idiom. Judges were instructed (see Appendix B) to rate each idiom along a 4 point scale, with 1 = very unfamiliar; 2 = somewhat uncommon; 3 = somewhat common; 4 = very familiar (see Levorato & Cacciari, 1992 for use of a similar familiarity scale). Instructions to each judge generally followed from Katz, Paivio, Marschark and Clark (1988). The idioms were

<sup>17</sup> The use of a single rater is contra-indicated. Reagan (1987), for example, found disagreements between  $N = 1$  ratings of familiar idioms published by Fraser (1970) and Newmeyer (1974).

presented to each judge in randomly ordered lists, to control for possible order effects. To probe for intra-observer rating stability, 16 of the idioms (10% of the total) in the main list (4 each of VP, NP, ADJP and ANOM, randomly selected from within each grouping) were re-presented to each judge as the final items of the main list. These 16 'probe' idioms were included in a fixed order as items #161 - 176. Per the directions given to each judge (Appendix B), the raters were instructed NOT to go back over the list and compare answers.

The judges' ratings of the familiarity of the idioms appeared stable. The mean rating for the 16 control idioms contained within the body of the randomized idiom list on first pass was 2.52; on re-presentation at the end of each list, the mean rating for the control idioms was 2.53. Item-to-item correspondence was 76.3%. Eighteen of the 19 changed scores (from  $N = 80$  for intra-rater reliability) were shifted by one scale point, one by two points.

A mean familiarity score (1 = unfamiliar; 4 = familiar) for each idiom was calculated from the ratings provided by the judges. Interestingly, of the 40 anomalous idioms presented, 38 were rated as relatively familiar (ratings over 3.0) by the five undergraduate judges. Indeed, it was not immediately obvious that adequately unfamiliar, anomalous idioms (examples include: *all in all*, *as for*, *catch as catch can*, and *every which way*) could be reliably identified using the selection methods employed here. A short list composed of an additional 13 potentially unfamiliar anomalous idioms was developed from Boatner and Gates (1966) by a second ASHA certified speech-language pathologist. Informal ratings by yet other SLPs suggested that these anomalous idioms were also too familiar for the design of the present study. Hence, the idiom form ANOMALOUS was excluded from the core multifactorial design. To provide descriptive and pilot data for further investigations, 8 anomalous idioms were selected and included in the experimental task. However, the data obtained were not included in the analyses for main effects, and are treated separately in the Results.

#### Stage 2 - Analyzability Ratings.

Next, five different undergraduates were paid to independently judge the degree to which the paraphrase of 89 less familiar idioms selected from the Stage 1 ratings could be figured out directly from the words comprising each

idiom. Each paraphrase of the idiom's figurative meaning was taken from Boatner and Gates (1966), or Makkai (1984).

Idioms for this pool were included based on the following:

- 1) Obtained mean familiarity rating for each idiom form below 2.4 (Mean familiarity rating across VP, NP, ADJP = 1.70, range = 1.58 – 1.90);
- 2) Include, based on experimenter ad hoc judgments and previous experimental work (Gibbs, 1991; Gibbs & Nayak, 1989), examples of both SA and SU idioms; and,
- 3) Include all 3 idiom forms. In addition, 10 randomly selected idioms were again presented as controls in a fixed order at the end of the main task to probe for intra-observer rating stability.

Judges were instructed (see Appendix A), following in general from Gibbs and Nayak's (1989) procedures, to rate each of the 89 idioms along a 5 point scale, with 1 = no relationship between the idiom's figurative meaning and the words that make up the idiom, 3 = a partial or indirect relationship between the idiom's figurative meaning and the words that make up the idiom (as in *bury the hatchet* meaning 'to resolve a dispute'), and 5 = a direct relationship between the idiom's figurative meaning and the words that make up the idiom. Again, the judges' ratings appeared stable: the mean rating for the controls on first pass was 3.70 ( $SD = 1.11$ ), on second pass 3.52 ( $SD = 1.05$ ). Calculation of a Pearson product-moment correlation coefficient of  $r = .96$ ,  $p < .05$ , was obtained between the two sets of ratings, indicating a high degree of agreement between the first- and second-pass ratings.

Stage 3 - Ecological Validity Check. All idioms selected were then discussed with three adolescent informants (nonparticipants). The purpose was to determine whether, despite the adults' judgments, any of the idioms proposed for use were seen as unusual in some fashion, or had a 'street' meaning unknown to the investigator. While each student was unfamiliar with one or more of the target idioms, none of the students reported anything unusual or remarkable per se about any of the items.

### Comprehension tasks

The idioms used in both comprehension tasks (item-selection and

explanation) were drawn from the pool of 89 idioms rated for analyzability by the five judges. A total of 8 idioms (4 SA, 4 SU) were selected for each of the idiom forms VP, NP and ADJP. The idioms used for each category (VP-SA, VP-SU; NP-SA, NP-SU; ADJP-SA, ADJP-SU) were chosen based on their mean familiarity rating, with the least familiar items selected first, combined with their SA - SU ratings, with the most and least analyzable idioms being selected for their respective categories. Table 6 contains the specific idioms selected, and Table 7, the idioms' mean familiarity ratings.

Table 6

Idioms used in the Comprehension Tasks

Idiom Form	<u>Analyzability</u>		<u>Idiom</u>	
	Analyzable	Unanalyzable		
Verb Phrase	X		Go to bed with the chickens	
	X		Meet your Waterloo	
	X		Lord it over	
	X		Ask for the moon	
		X	Carry coals to New Castle	
		X	Fly in the face of	
		X	Ring the changes	
		X	Give to understand	
	Noun Phrase	X		The horns of a dilemma
		X		A bee in her bonnet
X			The fat of the land	
X			A law unto one's self	
		X	A straw in the wind	
		X	A dressing down	
		X	A pig in a poke	
		X	The life of Riley	
Adjective Phrase	X		Behind the eight-ball	
	X		Out of the swim	
	X		Bred in the bone	
	X		Asleep at the switch	
		X	Beyond the pale	

Table 6 (continued)

Idiom Form	<u>Analyzability</u>		<u>Idiom</u>
	Analyzable	Unanalyzable	
Adjective Phrase (continued)		X	All wool and a yard wide
		X	Three sheets to the wind
		X	Fed to the teeth
Anomalous <sup>18</sup>			Come a cropper
			Tit for tat
			By and by
			By and large
			Hither and yon
			Which is which
			Less than no time
			Catch as catch can
Training	X	X	Get the bounce (VP) A case in point (NP)

Table 7

Mean Familiarity Ratings, by Idiom Form and Analyzability

<u>Rating</u>	<u>Form</u>			<u>Analyzability</u>	
	<u>VP</u>	<u>NP</u>	<u>ADJP</u>	<u>SA</u>	<u>SU</u>
Mean	1.53 <sup>1</sup>	1.85	1.73	1.72	1.68
SD	.32	.40	.52	.40	.46

<sup>1</sup> Unfamiliar = 1; Very familiar = 4

With the overall mean familiarity rating = 1.70, (SD = .43), this indicates that the idioms selected as stimuli were, overall, judged by the adult raters to be somewhere between 'very unfamiliar' and 'somewhat uncommon'.

<sup>18</sup> These Anomalous idioms were not balanced for familiarity (refer to Methods-Stimuli-Stage Two for additional information), or for analyzability. They are included for information purposes only.

To determine whether there were any significant differences in the levels of familiarity among the 3 idiom forms or between the categories SA - SU, separate ANOVAs were computed using the averaged familiarity ratings for each idiom. Results indicate that there are no significant differences between the SA - SU familiarity ratings,  $F(1,22) = .0353$ ,  $p = .8527$ , or between the familiarity ratings for the 3 Idiom Types,  $F(2, 21) = 1.2105$ ,  $p = .3180$ .

Eight Anomalous idioms (the 2 unfamiliar anomalous idioms and 6 others) were added to probe informally whether any gross differences were apparent between this form and the other idiom forms. Two additional idioms, one NP and one VP, were used to develop training items. Thus, a total of 34 idioms were selected for use in the comprehension tasks.

For each of the 34 idioms, short stories designed to clearly support an idiomatic interpretation of the idiom were developed. Each story consisted of two or three sentences plus a terminal sentence containing the target idiom as a sentence - final phrase<sup>19</sup>. Story development followed a general methodology first proposed by Ortony, et al., (1978)<sup>20</sup>. While absolute limits were not imposed, the stories were written to be generally balanced in length and informativeness. The topic of each story was designed to be of general interest and credibility to students across the G2 - G8 age range. Each story contained a character(s), setting and problem which combined to describe a short behavioral episode that would lead plausibly into the terminal sentence containing the target idiom<sup>21</sup>. Initial drafts of these stories were reviewed by a second speech/language pathologist for feedback regarding coherence, cohesion, plausibility, clarity and balance among the stories. After subsequent revision, the stories and item-selection questions (see below) were pilot-tested

<sup>19</sup> While there are no absolute guidelines set for story length, in practice most researchers have used stories totaling 3 - 5 sentences: Ezell and Goldstein, (1991): X = 24 words, 3 sentences; Nippold and Rudzinski, (1993): 4 sentence paragraph; Gibbs, (1991): 4 or 5 sentences, two actors having a conversation; Abkarian, (1992): 3 sentence story.

<sup>20</sup> Ortony, et al., (1978) used these procedures for developing stories to bias readers to (an appropriate) interpretations of stories: i) the context should clearly induce the figurative interpretation; ii) "the contexts themselves should be written using only literal language", and iii) "the target should not merely repeat or translate one of the context sentences but should be a continuation or summary sentence" (p.468)

<sup>21</sup> See also Levorato and Cacciari (1992, 1995) for additional discussion on the use of story grammar in developing plausible short stories that support the figurative meaning of each target idiom.

with three nonparticipant children (one each in G1, G4, and G5), and additional minor revisions were made using feedback from these students.

For each story, a set of three alternative paraphrases designed to probe comprehension of the idioms were developed as the item selection task. For the item-selection task, each set of three paraphrases included:

- a) the paraphrase of the idiom (e.g.: In the last sentence, does *kick the bucket* mean 'to die'), from Makkai (1984);
- b) the paraphrase of another idiom not otherwise used in the study (e.g.: In the last sentence, does *kick the bucket* mean 'give up'); and,
- c) a literal paraphrase (or close approximation) of the target idiom (e.g.: In the last sentence, does *kick the bucket* mean 'hit the bucket with your foot'?).

The use of three choices for the item-selection task follows generally from the multiple-choice comprehension task used by Levorato and Cacciari (1992, 1995), who found this task differentially sensitive through at least fourth grade. Inclusion of both explanation and item-selection tasks follows from recommendations by Nippold (Nippold & Rudzinski, 1993; Nippold & Taylor, 1995), who observed that each response task was differentially sensitive to identifying development profiles of figurative competence.

Order of presentation of the paraphrase types was counter-balanced across the target idioms to avoid the possibility that a participant would develop a response set to the order of choices presented. The idioms, short stories and associated paraphrases are in Appendix C.

### Procedures

All students with signed consent forms were administered the screening, selection, and experimental tasks. Students were tested individually in their own schools by the investigator (Glenn Johnson, M.A., CCC, S/LP). The screening and selection measures were administered in one session, the idiom comprehension tasks in a subsequent, follow-up session. Only the data from students who met the screening and selection requirements were included in the Results.

Participants were first instructed about the general nature of each experimental task, and then more specifically about what they were expected to

do during each segment (see Appendix D for instructions to the participants). Each session with each participant was audio taped for later analysis, to ensure procedural integrity in administration, and measurement reliability for scoring of the screening and experimental measures. The audiotapes were coded by Grade, Subject Number (from a separate master list to ensure confidentiality), and by session date. For example: G2S1:04/10/92 = Grade 2, Subject #01, April 10, 1992.

The experimental procedure was illustrated for each participant through the use of two training trials (Appendix C). A VP idiom was used on the first trial, an NP idiom on the second trial. Corrective feedback was provided as needed only during training. For both training and data collection, one story was read at a time. The examiner read each short story context and target idiom with normal prosody. Particular effort was made to read each target idiom without any prosodic cues that could bias a figurative interpretation of the idiom (see Gibbs, 1991, for an opposing view on prosody when speaking the target idiom).

After each story the child was first asked to explain what happened in the terminal sentence. If the child's response was incomplete or difficult to interpret, the investigator provided one prompt ('Tell me a little more'; or, 'I didn't quite get that, could you explain that some more?') in an attempt to elicit a more complete or interpretable response. Next, the student was read the target idiom, then the three item-selection paraphrases and asked to select the paraphrase that meant the same thing as the target idiom. One repetition of the story was allowed, at the student's request. Three randomly sorted lists of stimulus items were used (five participants at each age for each list), in order to control for possible order effects among the stimulus items.

During training and the first two experimental trials, participants were verbally reinforced for appropriate attending behaviors. General effort was reinforced as appropriate throughout the experimental tasks. For any questions pertaining to performance, the participant was reminded that the investigator could give no further information during that portion of testing. If a student indicated uncertainty about a response, or took more than 30 seconds to respond, the investigator encouraged the student to guess. All 'No responses' were recorded in writing.

### Response Measures

Comprehension tasks: The comprehension task had two response measures. The first response measure consisted of each participant's verbal explanation of the meaning of each idiom. The explanations were categorized into four classes (Correct Idiomatic; Correct Literal; Figurative-Other<sup>22</sup>; Unrelated), using a scoring system first developed by Ackerman (1982). The second response measure consisted of each participant's correct selection of a paraphrase of the idiom from the multiple choice comprehension measure. Here, each choice was categorized as either: (a) Correct-Idiomatic; (b) Literal; and, (c) Idiom-Other. This categorization allows for possible post hoc analysis of changes in preference for response categories. The response categories for each measure are outlined in Table 8.

### Reliability

Procedural reliability. Procedural reliability measures were conducted for two separate areas: (1) for administration and scoring the standardized tests used for participant selection; and (2) for administration of the experimental measures (the comprehension tasks). The outcomes of these reliability probes are reported next.

Procedural reliability for administration and scoring of the tests used for participant selection was completed in two steps. First, an ASHA certified clinician who did not participate in participant selection reviewed audiotapes of 9% of the administered tasks, randomly selected, for compliance with standardized test administration procedures. Overall compliance with standardized administration procedures was 97% (obtained by comparing the actual procedures performed by the investigator with the total number of procedures expected per the published administration procedures<sup>23</sup>). Next, the accuracy of the scoring for the standardized test measures was determined by having the second clinician review scored protocols from 9% of participants for

<sup>22</sup> For additional discussion of, and rationale for the use of this four-part categorical scoring, see Levorato and Cacciari (1995), particularly pp. 263 - 266, for a discussion of their development of a stage model of figurative competence. In this model the use of 'figurative-other' corresponds to their 'figurative' stage. This is an intermediate developmental step, in which children indicate their realization of the nonliteralness of the idiom, but cannot provide an accurate idiomatic answer.

<sup>23</sup> The general procedure follows from Billingsley, White and Munson (1980).

Table 8

Response Categories and Criteria for the Explanation and Item-selection Comprehension Tasks

Comprehension Task	Response Category	Criteria
Explanation	1. Correct-Idiomatic	The student's explanation follows a synthesis of idiom dictionary definitions <sup>24</sup> and/or paraphrases the idiom's meaning within the story.
	2. Correct-Literal	The student explains the correct literal meaning of the idiomatic phrase. All verbatim repetitions of the target idiom are assigned to this category.
	3. Figurative-Other	The student's explanation conveys some knowledge that the meaning of the idiom is not just made up of the words comprising the idiom, <b>BUT</b> the explanation does not accurately capture the dictionary definition or a plausibly accurate paraphrase of the idiom's figurative meaning. Explanations that convey thematic or pragmatic essence of the story without accurately conveying the idiom's figurative meaning are assigned to this category.
	4. Unrelated	The student's explanation does not reflect either the idiom in any sense, or the essence of the story. 'No responses' are assigned to this category.

<sup>24</sup> The recognition rules for each idiom are provided in Appendix D.

Table 8 (continued)

Item-selection	1. Correct-Idiomatic	Selects the correct figurative paraphrase from the list of three choices.
	2. Correct-Literal	Selects the correct literal paraphrase from the list of three choices.
	3. Idiomatic-Other	Selects the distracter idiom from the list of three choices.

---

compliance with standardized test scoring protocols. Overall compliance with scoring protocols was 98% (based on item-by-item scoring agreement among the investigator and reliability clinician).

Procedural reliability for administration of the experimental tasks was obtained by having a speech/language graduate clinician-in-training and an ASHA certified clinician review 13% of the audiotapes (2 each @G2, G5 & G8), randomly selected, for compliance with key points in the procedures outlined in Procedures above. The independent reviewers evaluated each tape for: (i) the use of normal prosody in the reading of the stories and target idioms, using a two-point scale (1 = acceptable, neutral-to-natural prosody; 2 = exaggerated prosody, over-reading the material); and, (ii) general (i.e., 'good listening') versus explicit (e.g.: 'yes, that's right') verbal reinforcement of responses. Procedural integrity in these two areas was judged as critical to ensuring that participants' responses were not unintentionally biased by the differential skewing of meanings (through prosodic cuing) or in the differential reinforcement of responses.

Averaged ratings by the two judges for prosodic cuing, and for the appropriate use of reinforcement are presented in Table 9.

Compliance with the protocols for the experimental tasks was satisfactory for both prosody and reinforcement for G5 and G8. Compliance with use of normal prosody, and in the use of general verbal reinforcement was somewhat lower for G2. In retrospect, this was likely due to a tendency to 'sell' and 'support' (with praise) the tasks as interesting to this younger group, and by

unconsciously using a 'story time' prosody in doing so. Since order of administration of the stimulus items was randomized, no within-group effect is expected.

Table 9

Procedural Reliability: Judges Averaged Ratings for Prosody and Use of Verbal Reinforcement.

Grade	Rating	
	Prosody <sup>1</sup>	Reinforcement <sup>2</sup>
2	1.07	1.08
5	1.05	1.02
8	1.03	1.03

<sup>1</sup> 1 = normal, neutral prosody; 2 = exaggerated prosody. Lower average ratings indicate compliance with the outlined procedures.

<sup>2</sup> 1 = either appropriate use of neutral reinforcement, or no reinforcement; 2 = explicit verbal reinforcement of responses. Lower average ratings indicate compliance with the outlined procedures.

Measurement reliability. Measurement reliability was conducted for both the transcribing of the audiotapes and for the categorizing of the participants' verbal explanations in the explanation task. Participants' responses in the explanation task were first orthographically transcribed. Next, for transcription reliability, 13% (two each at G2, G5 & G8) of the participants' audiotapes were then randomly selected to assess interobserver reliability. Each tape was reviewed by a speech/language graduate clinician-in-training for agreement with the transcribed explanations. Point-to-point transcription agreement was 94.7% for content words and connectives. Subsequent mutually-agreed upon changes in the transcript changed the initial scoring by only .5% (.005).

For reliability in categorizing the participants' verbal explanations of the figurative meaning of each idiom, 25% of the participants' audiotapes (three

each @G2, G5 & G8) were randomly selected to assess interobserver reliability. The principal investigator first reviewed and made an initial categorization of each response, using the four response categories outlined in Table 8, above, and using definitions for acceptable figurative paraphrases derived from idiom dictionaries by Makkai (1987) (primary source), and Spears (1995) (secondary source; used to clarify some paraphrases as needed). The dictionary definitions were used to create a set of recognition rules for identifying the correct figurative paraphrase of each stimulus item. The recognition rules contained: (i) the correct figurative paraphrase; and (ii) an example of the idiom used figuratively in a sentence that was different from the sentence used in the explanation tasks.

Next, a speech/language graduate clinician-in-training independently reviewed and categorized the participants' transcribed explanations, using the same recognition rules. Point-to-point agreement was 81.6%, which was deemed unacceptable. Subsequently, a more explicit set of recognition rules for categorizing the explanations were developed, reviewed by a second ASHA certified clinician for sensibility and explicitness, and used by both the principal investigator and the graduate reliability observer to categorize participants explanations again (25% of audiotapes, 3 each at G2, G5 & G8).

Point-to-point agreement was 91.3%. The responses for which the judges disagreed were subsequently reviewed, discussed, and categorized again as needed by mutual consent. For target idioms where the review of the disagreements in categorizing led to yet further explication of the recognition rules, all participant responses were categorized again in agreement with the revised rules.

### Design

The experimental design is a 3 (Grade Level - G2, G5 & G8) by 3 (Idiom Form-VP, NP & ADJP) x 2 (Analyzability-Semantically Analyzable & Semantically Unanalyzable) mixed factorial design. The factors Idiom Form and Analyzability are treated as within-subjects repeated measures. The data obtained consisted of: (1) the item choices made by each participant at each grade level in the item-selection comprehension task; and, (2) the categorized explanations given by each participant when asked the meaning of each target

idiom in short stories (see Table 8 for categories and scoring criteria).

### Data Analysis

The data obtained were analyzed with respect to the questions posed in the Statement of the Problem. Within each Question, analysis of variance (ANOVA) tests for main effects were calculated, followed by post hoc statistical analyses when appropriate. The data from the item-selection and explanation tasks were analyzed separately. Each set of responses was analyzed for significant main effects and interactions among significant main effects using 3 (Grade) X 3 (Idiom Form) X 2 (Semantic Analyzability) multivariate ANOVAs (MANOVA). An alpha level of  $p \leq .05$  was used throughout. When significant ( $p \leq .05$ ) main effects were found, post-hoc tests using Scheffe's procedures (see Kerlinger, 1971) were used to determine which mean differences, if any, contribute significantly ( $p \leq .05$ ) to the overall variance.

The research questions addressed are as follows:

1) Does the analyzability (SA - SU) of idioms influence children's comprehension of idiomatic meanings?

2) If analyzability differentially affects the acquisition of figurative competence for idioms at any grade, are there any Analyzability x Grade interactions?

For Questions 1 and 2, the independent variable is grade (G2 - G8), and the dependent variable is analyzability (SA - SU). Univariate ANOVAs will test for differences among SA and SU idioms, for the growth of comprehension competence with each increase in grade, and for any Analyzability x Grade interactions.

3) Does the syntactic form (VP, NP, ADJP) of an idiom influence children's comprehension of idiomatic meanings? And, more specifically:

4) If a main effect for idiom form is obtained, how does the developmental course of figurative competence for noun phrase and adjective phrase idioms compare to the developmental course for verb phrase idioms across G2, G5

and G8?

5) If the developmental course of idiom comprehension differs based on idiom form, does the effect of form vary based on grade?

For Questions 3, 4 and 5, the dependent variable is idiom form (3), and the independent variable is grade (3). A univariate ANOVA was used to test for significant differences in the acquisition of idiomatic competence among VP, NP, and ADJP idioms, for the growth of comprehension with each increase in grade, and for any two-way Form x Grade interactions.

6) Are there interactions between analyzability and idiom form in children's acquisition of figurative meanings?

7) Are there any three-way interaction effects among analyzability, idiom form, and grade in children's acquisition of idiomatic meanings?

For Questions 6 and 7, the dependent variables are analyzability (SA - SU) and idiom form (VP, NP, & ADJP), and the independent variable is grade (G2, G5, & G8). A multivariate analysis of variance (MANOVA) was conducted to test for possible interactions among analyzability x form, and for any possible three-way interactions among analyzability, form and grade.

8) How does children's acquisition of ill-formed or anomalous (ANOM) idioms compare with their acquisition of well-formed idioms?

Descriptive data were provided to detail the development of figurative competence in anomalous idioms.

All the statistical tests were computed using the SPSS for Macintosh software package (release 6.1.1; 1996).

## CHAPTER THREE

### Results

The experimental results are reported with reference to each of the eight questions posed in the Statement of the Problem. For questions #1-7, the results of repeated measures univariate and multivariate analyses of variance (ANOVA, MANOVA), which seek to determine whether significant main effects can be identified for the two dependent (analyzability and idiom form) and one independent (grade) variables, are presented first. These are followed by post hoc statistical and descriptive analyses when appropriate. A significance level of  $p \leq .05$  is used throughout.

The experimental results for Question #8 are reported and treated differently. Question #8 focuses on the developmental data for Anomalous Idioms (ANOM). As originally detailed in 'Methods', ANOM idioms are not matched with the other idiom forms on either of the dependent measures (analyzability, familiarity). Given the critical differences in these stimulus parameters, the data from the ANOM idioms cannot reasonably be included in the planned comparisons that form the basis for the ANOVAs reported below. Thus, only descriptive data are provided to detail the development of figurative competence in ANOM idioms.

Initially, separate analyses were completed for the data from the item-selection and explanation tasks respectively. This was done because, as Nippold argues (Nippold & Rudzinski, 1993; Nippold & Taylor, 1995), these tasks tap overlapping but distinct elements of idiomatic competence. Each analysis is presented sequentially (item-selection, followed by explanation) within each question.

As the results from Questions #1-7 were evaluated and integrated in the Discussion, it became clear that an additional (and unforeseen) interaction, between the two dependent variables and the response tasks, had occurred. To test this proposed interaction statistically, the response tasks were subsequently treated as a dependent variable (task), and an additional MANOVA was completed using three dependent variables (Analyzability x Form x Task) and

and one independent variables (grade). The results of this MANOVA are presented in the section Task Interactions; the reasoning which supports the use of this statistical test is contained in the Discussion.

A total of 1728 response opportunities, 864 from each response task (= [4 SA + 4 SU] x 3 idiom forms x 12 participants x 3 grades) were used in the statistical and descriptive analyses relevant to Questions #1-7, and for investigating the 'task interactions.. For Question #8, which focuses on the ANOM idioms, a total of 576 responses, 288 (= 8 ANOM x 12 participants x 3 grades) for each of the item-selection and explanation tasks, are used in the presentation of the descriptive analyses.

The statistical and descriptive data related to each of the 8 research questions posed in Chapter 2 are presented next.

Question 1. Does the analyzability (SA - SU) of idioms influence children's comprehension of idiomatic meanings?

For Questions #1 and 2, the dependent variable is analyzability (SA - SU), and the independent variable is grade (G2 - G8).

Item-selection response task.

Mean percentage of correct responses for analyzable (SA) and unanalyzable (SU) idioms at each grade were calculated (denominator = 144 for each cell; range in the numerator = 88 - 137), and the results presented in Table 10. The results summarized in Table 10 indicate that students were more successful in selecting the meanings of SU idioms ( $M = 86.1\%$  accurate), than of SA idioms ( $M = 71.1\%$  accurate), with comprehension of both SU and SA idioms increasing with each increase in grade level. Inspection of the results in Table 10 also indicates that not only did participants in each grade select more correct SU responses than SA responses, but by G8 the students' comprehension of SU idioms had effectively reached a performance ceiling. Gibbs (1991) reported similar ceiling effects in item-selection tasks as early as G3 when familiar idioms are used as stimuli.

To determine the statistical significance of these observations, a 2 (Analyzability) x 3 (Grade) mixed ANOVA (ANOVA1), with repeated measures on the first factor, was performed. ANOVA1 (Table 11) yielded significant main

effects for the factor analyzability, with more correct responses for SU than for SA idioms, and for grade. That is, not only did analyzability not aid students' performance, as would be predicted from Gibbs' (1991) findings and his general arguments regarding analyzability, but students' comprehension of analyzable idioms was actually significantly worse than their comprehension of unanalyzable idioms.

Table 10

Mean Number and Mean Percentage of Correct<sup>1</sup> Item-Selection Responses for SA and SU Idioms, by Grade, with Post Hoc Significant Differences for Grade

Grade	<u>Analyzability</u>						<u>Mean</u> <u>% correct</u>
	<u>Analyzable (SA)</u>			<u>Unanalyzable (SU)</u>			
	<u>M</u>	<u>SD</u>	<u>% correct</u>	<u>M</u>	<u>SD</u>	<u>% correct</u>	
2	7.33*	2.64	61.1	9.17*	2.08	76.4	68.8
5	8.25	1.48	68.8	10.42	1.08	86.8	77.8
8	10*	1.28	83.3	11.42*	.79	95.1	89.2
Mean	8.53	2.16	71.1	10.33	1.62	86.1	78.6

Note. Maximum score = 12

<sup>1</sup> chance = 33.3% correct

\* differences in correct responses were significant between these grades,  $p = .05$ , per post hoc testing using Scheffe's procedures.

To determine between which grades significant differences in selection of the idiomatic paraphrase occurred, a post hoc analysis of the results was completed with a one-way ANOVA using Scheffe's procedures for multiple comparisons (Table 10). This analysis was completed in order to more fully characterize the growth of figurative competence across the grade range tested.

Scheffe's procedure was used because, per Ferguson (1981), the "method is more rigorous than other multiple comparison methods with regard

to Type I error [producing fewer significant differences]...[and] it is not seriously affected by violations of the assumptions of normality and homogeneity of variance..." (p. 308 - 9). Because one goal of this dissertation is to help clarify conflicts in an area of research (analyzability) that already contains at least a modest body of research, the use of a statistical method which minimizes 'false positive' findings is desirable. Further, while considerable effort was used to ensure that the participants are developmentally typical on important cognitive and linguistic parameters, it is not possible to control all possible developmental parameters. For example, participants' exposure to figurative language in the home, which may vary to some degree across families, cannot be controlled for directly, and participants may, therefore, have more or less practice in glossing figures of speech. Hence, use of a statistic that 'holds up well' to potential violations of assumptions of homogeneity also seems desirable, and this contributed to the selection of Scheffe's procedure.

Table 11

Analysis of Variance for Students' Item-Selections of SA and SU Idioms (ANOVA1).

Source	E	
	df	SA - SU
Between subjects		
Grade (G)	2	8.63*
<u>S</u> within-group error	33	(1.41)
Within subjects		
Analyzability (SA)	1	41.09**
G x SA	2	.59
SA x <u>S</u> within-group error	33	(.48)

Note. Values enclosed in parentheses represent mean square errors. S = subjects.

\* $p = .001$       \*\* $p < .000$

This post hoc analysis showed significant increases ( $p = .05$ ) in correct responses between G2 and G8 for both SA and SU idioms. No other between-grade differences were significant. That there were no significant differences in performance between the G2 and G5 students is perhaps not surprising, given the relatively high levels of accuracy that the G2 students demonstrated in selection of the figurative paraphrase in both conditions (Table 10).

#### Explanation Task.

Mean percentage of correct responses for analyzable and unanalyzable idioms at each grade was calculated (denominator = 144 for each cell; range in the numerator = 18 - 63), and these results are presented in Table 12. These Table 12

#### Mean Number, and Mean Percentage of Correct Explanations of SA and SU Idioms, by Grade, with Scheffe` post hoc Significant Differences for Grade.

Grade	<u>Analyzability</u>						<u>Mean</u> <u>% correct</u>
	<u>Analyzable (SA)</u>			<u>Unanalyzable (SU)</u>			
	<u>M</u>	<u>SD</u>	<u>% correct</u>	<u>M</u>	<u>SD</u>	<u>% correct</u>	
2	1.5*#	1.62	12.5	1.67*	1.72	13.2	12.9
5	4.25#	1.29	32.6	2.83	1.19	25.7	29.2
8	5.33*	1.5	43.8	4.33*	2.02	36.8	40.3
Mean	3.69	2.18	29.6	2.94	1.97	25.2	27.4

Note. Maximum score = 12

\* differences in correct responses were significant between G2 - G8,  $p = .05$ , per post hoc testing using Scheffe's procedures

# differences in correct responses were significant between G2 - G5,  $p = .05$ , per post hoc testing using Scheffe's procedures

results show, in contrast with the outcomes from the item-selection task, which favored SU>SA idioms, that students in both G5 and G8 were more successful in explaining the meanings of SA idioms than of SU idioms. However, for G2 students, who found this explanation task quite challenging (providing, on

average, one correct explanation for each eight items given), there was no real difference in accuracy between SA and SU idioms. Students' ability to explain the figurative meanings of both SA and SU idioms increased with each increase in grade, though this task remained challenging for the G8 students.

To investigate the statistical significance of these observations, a 2 (Analyzability) x 3 (Grade) mixed ANOVA (ANOVA2), with repeated measures on the first factor, was performed. ANOVA2 (Table 13) yielded a significant main effect for grade, but no significant main effect for analyzability. The absence of a facilitative effect for analyzability in the verbal explanation task is similar to one outcome reported by Nippold (1993), who found that, in a written explanation task, no effect for analyzability was identified in G5 students' explanations (though positive effects for analyzability were reported at G8 and G11).

Table 13

Analysis of Variance for Students' Explanations of SA and SU Idioms (ANOVA2).

E		
Source	df	SA - SU
Between subjects		
Grade (G)	2	19.94*
<u>S</u> within-group error	33	(1.41)
Within subjects		
Analyzability (SA)	1	3.92
G x SA	2	1.30
SA x <u>S</u> within-group error	33	(.47)

Note. Values in parentheses represent mean square errors. S = subjects.

\* $p < .000$

To determine between which grades significant differences in explanations of the idiomatic paraphrase occurred, a post hoc analysis of the results was completed with a one-way ANOVA using Scheffe's procedures for multiple comparisons, and the results included in Table 12. This analysis was completed in order to more fully characterize the growth of figurative competence across the grade range tested.

This post hoc analysis showed significant differences in correct explanations between G2 and G5, and between G2 and G8 for SA idioms, and between G2 and G8 for SU idioms. No other between-grade differences were significant.

While no significant effect for analyzability was found in the present research, a moderate trend favoring more correct explanations for SA over SU idioms in G5 and G8 is apparent from inspection of Table 12. This trend was investigated further through two supplemental, a posteriori ANOVAs, reported below.

The G2 students obviously had substantial difficulties in correctly explaining the meanings of these unfamiliar idioms, difficulties not shared by the G5 and G8 students. The  $M = 12.9\%$  correct result for the G2 students means that, on average, each participant produced 4 accurate and 28 inaccurate or otherwise incomplete explanations. Yet, these same G2 students accurately selected many correct meanings in the item-selection condition ( $M = 68.8\%$  accurate). This difference in performance (68.8% correct item-selection  $\nu$  12.9% correct idiomatic explanations) suggests that the G2 students found the metalinguistic demands of the explanation task to be quite high. Such a conclusion and finding is perhaps not surprising since, as reported earlier in Studies of Typical Development, students in several other studies have been similarly challenged by explanation tasks (Levorato & Cacciari, 1995; Prinz, 1983). Thus, it is possible that the 'high bar' presented by this specific task may in some sense override any potential differences in the contribution of SA  $\nu$  SU idioms, at least when the G2 students' performance is included.

To explore this hypothesis, two additional, a posteriori ANOVAs were performed. For the first additional ANOVA (ANOVA3), the students' Figurative-

Other' (F-O) responses<sup>99</sup> were added to their 'Correct-Idiomatic' answers, and the combined scores (Correct-Idiomatic + Figurative-Other) used to explore the contribution of SA - SU to figurative competence. The results of the 2 (Analyzability) x 3 (Grade) ANOVA3 (Table 14) again determined that there were no significant differences between SA and SU idioms in students' demonstration of figurative competence in the explanation task.

Table 14

Analysis of Variance for Students' C-I + F-O Explanations of SA and SU Idioms (ANOVA3).

Source	E	
	<u>df</u>	SA - SU
	Between subjects	
Grade (G)	2	32.6*
<u>S</u> within-group error	33	(1.58)
	Within subjects	
Analyzability (SA)	1	3.48
G x SA	2	1.16
SA x <u>S</u> within-group error	33	(.53)

Note. Values enclosed in parentheses represent mean square errors. S = subjects.

\*p < .000

<sup>99</sup> see Table 5. These are explanations which indicate an awareness that the meaning is nonliteral, but which do not accurately convey the figurative meaning of the target idiom. Examining these responses along with the Correct - Idiomatic responses creates an alternative mechanism for identifying whether the students are at least 'heading in the right direction', or 'talking toward' the correct figurative meanings in their attempts to interpret the idioms to the examiner. See Levorato and Cacciari (1995) for additional discussion.

To further describe and explore the students' abilities to at least 'talk toward' the correct idiomatic meanings, the mean percentage of F-O responses for analyzable and unanalyzable idioms at each grade was also calculated (denominator = 144 for each cell; range in the numerator = 43 - 64). These results are presented with the 'Correct-Idiomatic' (C-I) explanations, and with their combined (C-I + F-O) totals, in Table 15. The results summarized in Table 15 show that students in every grade were working strongly toward a figurative interpretation (range = 42.4 - 85.4% figurative), as examination of the students' responses in the item-selection task (Table 10) has already suggested. However, these (C-I + F-O) descriptive data only reinforce the results from the ANOVA3 reported above, of no significant difference between SA - SU idioms for the three grades, even when all the figurative explanations are considered.

Table 15

Mean percentage of Correct - Idiomatic (C-I), Figurative - Other (F-O), and C-I + F-O explanations of SA and SU Idioms, by Grade.

Explanation Category	Analyzability					
	Analyzable (SA)			Unanalyzable (SU)		
	C-I	F-O	C-I + F-O <sup>1</sup>	C-I	F-O	C-I + F-O <sup>1</sup>
Grade						
2	12.5	29.9	42.4	13.2	29.9	43.1
5	32.6*	44.4	77.1	25.7*	43.8	69.4
8	43.8*	41.7	85.4	36.8*	43.1	79.9
Mean	29.6	38.7	68.3	25.2	38.9	64.1

<sup>1</sup> C-I + F-O totals may differ due to rounding

\* differences in correct C-I responses were significant between G5 - G8,  $p = .05$ , per ANOVA4

Next, the G2 students' responses were dropped from the data, and the G5 and G8 students' explanations were reanalyzed. The G2 students' responses were dropped because, as argued above, their performance in the explanation task suggests that the task demands overwhelmed the G2 participants' opportunity to demonstrate an effective analyzability strategy for figurative meanings. If this is so, then the results from the G2 participants may in effect 'mask' the use of such a strategy by the older students, who demonstrated greater facility with the metalinguistic demands of the explanation task.

The 'Correct - Literal' responses from the G5 and G8 students were analyzed using a 2 (Analyzability) x 2 (Grade) repeated measures ANOVA (ANOVA4), with joint univariate Scheffe' confidence intervals set at .9500 for each contrast. Scheffe's procedures were again used because: (i) this is an appropriately conservative statistic, as argued above, which (ii) controls well for experiment-wise error rates (Ferguson, 1981). The results of this analysis yielded a significant effect for analyzability,  $F(1, 22) = 6.26, p = .020$ , with G5 and G8 students correctly explaining the figurative meanings of more SA than SU idioms; and for grade  $F(1, 22) = 6.69, p = .017$ , with G8 students producing more correct explanations than G5 students.

Question 2. If Analyzability differentially affects the acquisition of figurative competence for idioms at any grade, are there any interactions between analyzability and grade?

Item-selection response task.

The effect of analyzability did not vary based on grade in the item-selection response task. Results of ANOVA1 (Table 11) determined that there were no significant interactions between analyzability and grade. The students consistently responded more accurately to SU than to SA idioms at each grade level.

Explanation task.

No statistically significant Analyzability x Grade interaction effects were identified from ANOVA2, ANOVA3 or ANOVA4. A total of three univariate ANOVAs were computed to examine the two-way interactions between Analyzability x Grade : In ANOVA2 for main effects (Question 1), and in each of

the two supplemental ANOVAs (for combined C-I + F-O, and for G5 + G8 students only). Results of ANOVA2 completed for Question 1 determined that there were no significant interactions between analyzability and grade. While the G2 students did produce slightly more correct SU than SA explanations ( $\underline{M}$  SU = 13.5%;  $\underline{M}$  SA = 12.9% correct), this finding is statistically trivial.

Results from ANOVA3, which used combined C-I + F-O explanations to test for significance, also found no significant Analyzability x Grade interactions. Finally, in ANOVA4, which examined for SA - SU main effects for G5 and G8 participants only, no significant Analyzability x Grade interactions were identified for the two grades used in the analysis,  $F(1, 22) = .01$ ,  $\underline{MSE} = .01$ ,  $p = .906$ , indicating that the effect of analyzability did not vary as a function of grade.

In summary, the analyses presented for Questions 1 and 2 determined:

- a) there was a significant main effect for analyzability across grades in the item-selection task; but,
- b) functionally, the impact of analyzability on comprehension was negative - i.e., students' selected more correct unanalyzable than analyzable idioms (SU>SA);
- c) there was no significant main effect for analyzability in the explanation task, due in part to the considerable difficulties that the G2 students had in explaining the figurative meanings; however,
- d) a significant, and facilitating, effect for analyzability (SA > SU) was identified when only the G5 and G8 students' responses were considered in the explanation task.

These results, considered together, indicate that an interaction effect occurs between analyzability and response task- i.e., item-selection x analyzability = SU>SA; explanation x analyzability = SA>SU. This unexpected finding will be treated further in the discussion.

Next, the results relevant to answering the research questions posed with regard to the impact of an idiom's syntactic form on idiom comprehension are presented. More generally, these analyses were performed to determine whether the developmental evidence on idiom comprehension derived from the nearly exclusive use of VP idioms does in fact generalize to other frequently occurring idiom forms, specifically NP and ADJP idioms.

Question 3. Does the syntactic form (VP, NP, ADJP) of an idiom influence children's comprehension of idiomatic meanings?

For Questions #3, 4 and 5, the dependent variable is idiom form (3), and the independent variable is grade (3). A joint univariate ANOVA was used to test for significant differences in the acquisition of idiomatic competence among VP, NP, and ADJP idioms, for the growth of comprehension with each increase in grade, and for any two-way Form x Grade interactions.

Item-selection response task.

Mean percentage of correct responses for VP, NP and ADJP idioms at each grade was calculated (denominator = 96 for each cell; range in the numerator = 65 - 88), and the results are presented in Table 16. Inspection of the results summarized in Table 16 indicates that an idiom's syntactic form does not affect comprehension of figurative meanings in the item-selection task by students in G2 - 8. Participants were essentially equally successful in selecting the meanings of VP, NP, and ADJP idioms. Even the youngest (G2) students demonstrated strong comprehension of figurative meanings in each idiom form, with little variation in success across forms at that grade (range of Ms = 67.8-70.9% correct). As inspection of Table 16 also indicates, the task of selecting the idiomatic meaning for each idiom was largely complete by G8.

Table 16

Mean Percentage of Correct Item-Selection Responses for each Idiom Form, by Grade.

Grade	<u>Idiom Form</u>								
	<u>Verb Phrase</u>			<u>Noun Phrase</u>			<u>Adjective Phrase</u>		
	<u>M</u>	<u>SD</u>	<u>% correct</u>	<u>M</u>	<u>SD</u>	<u>% correct</u>	<u>M</u>	<u>SD</u>	<u>%correct</u>
2	5.42	1.93	67.8	5.67	1.16	70.9	5.42	1.93	67.8
5	6.08	1.0	76.0	6.17	.94	77.1	6.42	.9	80.0
8	7.25	.62	90.6	6.83	.72	85.4	7.33	.65	91.6
Mean	6.25	1.48	78.1	6.22	1.05	77.8	6.39	1.48	79.9

To confirm that no statistically significant differences in comprehension exist between each idiom form, a 3 (Form) x 3 (Grade) mixed ANOVA (ANOVA5), with repeated measures on the first factor, was conducted. As expected from inspection of the data in Table 16, no significant main effect was obtained for idiom form,  $F(2, 66) = .40$ ,  $p = .674$ , in the item-selection task.

Explanation task.

Mean percentage of correct responses for VP, NP and ADJP idioms at each grade were calculated (denominator = 96 for each cell; range in the numerator = 7 - 51), and the results presented in Table 17. Inspection of the results summarized in Table 17 suggests that syntactic form has at least a modest effect on participants' success in explaining the figurative meanings of the target idioms. Students in each grade were always most successful in explaining the figurative meanings of VP idioms ( $M = 35.1\%$  accurate), and least successful in explaining the meanings of NP idioms ( $M = 21.6\%$ ). In each grade the accuracy of explanations for ADJP idioms occupied a middle ground. Percentage of correct explanations increased for each form with each increase in grade level. The results contained in Table 17 also indicate that, for this obviously challenging task, only the G8 participants' explanations of VP idioms exceeded the *break even* point of 50% accurate ( $M = 53.1\%$  accurate).

A 3 (Form) x 3 (Grade) mixed ANOVA (ANOVA6), with repeated measures on the first factor, was conducted to determine whether there were statistically significant differences between participants' explanations of each idiom form. ANOVA6 yielded a significant main effect for idiom form,  $F(2, 66) = 6.35$ ,  $p = .003$ . This indicates that the syntactic form of an idiom does have some differential impact on participants' success in explaining figurative meanings.

Next, dependent measures t-tests were performed on the data in order to identify among which idiom forms significant differences in the growth of idiomatic comprehension occurred. This analysis was done because another goal of this dissertation was to determine if, and to what degree, the developmental data derived primarily from VP idioms can be generalized to other idioms forms. These analyses address this goal, and their results are presented under Question #4, next.

Table 17

Mean Number and Mean Percentage of Correct Explanations for VP, NP and ADJP Idioms, by Grade.

Grade	Idiom Form								
	<u>Verb Phrase</u>			<u>Noun Phrase</u>			<u>Adjective Phrase</u>		
	<u>M</u>	<u>SD</u>	<u>% correct</u>	<u>M</u>	<u>SD</u>	<u>% correct</u>	<u>M</u>	<u>SD</u>	<u>%correct</u>
2	1.33*	1.37	16.7	.58 <sup>*</sup>	.79	7.3	1.17*	1.19	12.9
5	2.83 <sup>#</sup>	1.59	35.4	1.83 <sup>^</sup>	.84	22.9	2.42	1.38	29.2
8	4.25 <sup>#</sup>	2.01	53.1	2.67*	1.56	33.3	2.75*	1.36	40.3
Mean	2.81	2.03	35.1	1.69	1.39	21.2	2.11	1.45	27.5

\* differences in correct responses were significant between G2 - G8,  $p = .05$ , per post hoc testing using Scheffe's procedures

# differences in correct responses were significant between G5 - G8,  $p = .05$ , per post hoc testing using Scheffe's procedures

<sup>^</sup> differences in correct responses were significant between G2 - G5,  $p = .05$ , per post hoc testing using Scheffe's procedures

Question 4. If a main effect for idiom form is obtained, how does the developmental course of figurative competence for NP and ADJP idioms compare to the developmental course for VP idioms across G2, G5 and G8?

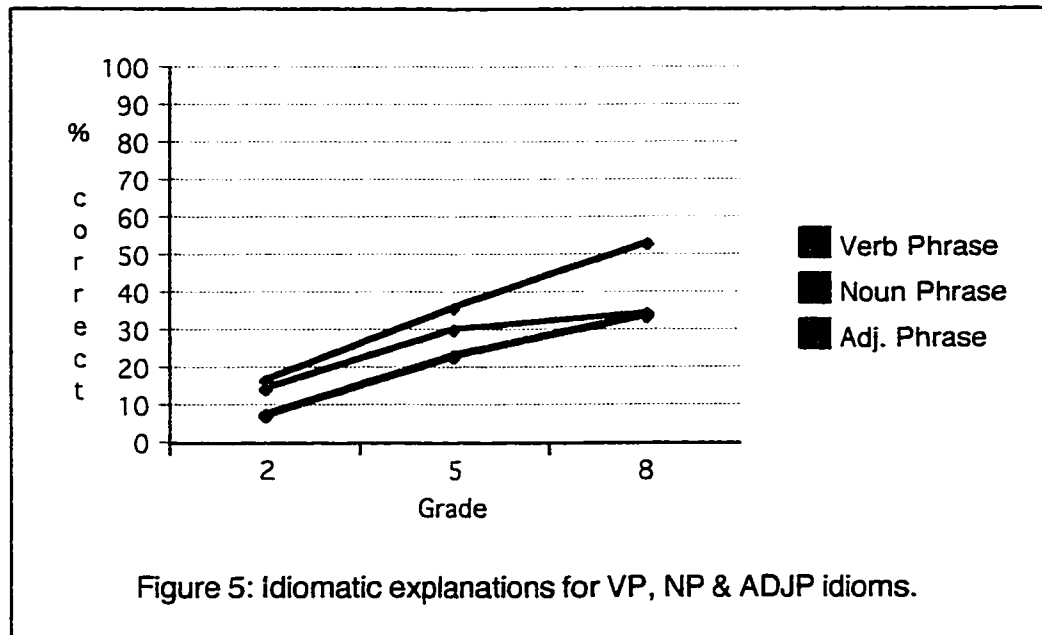
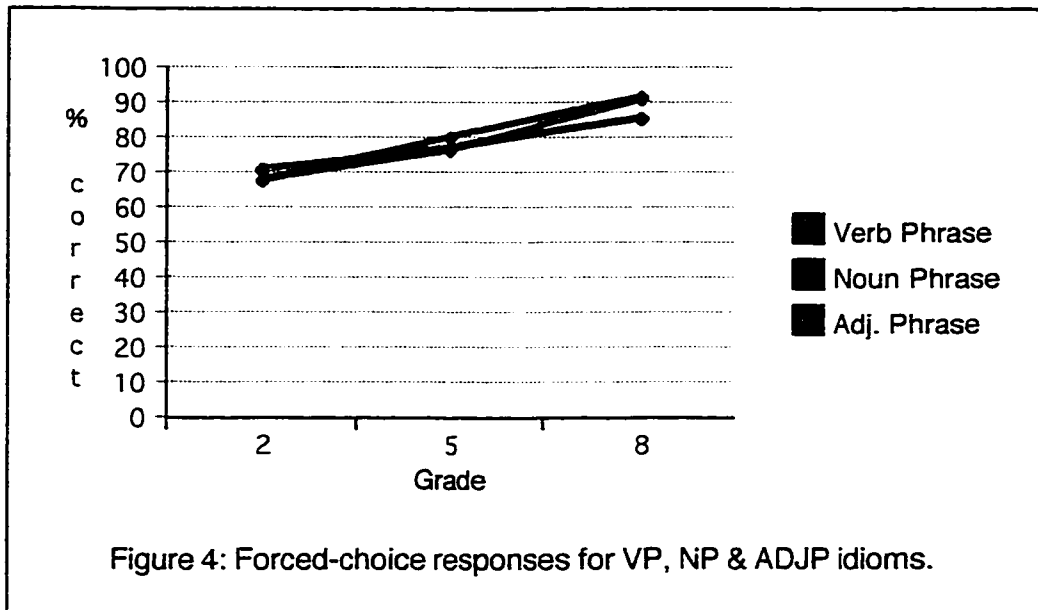
Item-selection response task.

As reported in Question #3, no main effect was found for idiom form in the item-selection response task. Represented graphically in Figure 4, the developmental course of acquisition of idiomatic meanings, as measured by the item-selection task, is quite similar for VP, NP and ADJP idioms at each grade.

Explanation task.

Supplementary and post hoc analyses of the results reported in Question 3 (Table 17), and presented graphically in Figure 5, were completed to determine: (a) between which idiom forms significant differences occur; and

(b) between what grades significant differences in correct explanations occurred for each idiom.



Visual inspection of Figure 5 suggests that there are consistent differences in the development of figurative competence for VP and NP idioms, and modest differences between VP and ADJP idioms. To examine these apparent differences statistically, a series of dependent measures  $t$ -tests was conducted for each pair of idiom forms: VP x NP; VP x ADJP; and NP x ADJP idioms. The Bonferroni procedure was used for this analysis in order to: (i) correct for the increased possibility of Type-I errors with multiple dependent  $t$ -tests; and (ii) because the Bonferroni procedure provides more power than Scheffe's procedure in identifying significant differences between means, while continuing to control well for Type-I errors (O'Brien & Kaiser, 1985).

Significant differences were obtained in the number of correct explanations between VP and NP idioms,  $t = 3.42311$ ,  $p = .0017$ ; and between VP and ADJP idioms,  $t = 2.07095$ ,  $p = .0463$ , with students producing more correct explanations for the VP idioms than for either NP or ADJP idioms. However, no significant differences were obtained in the number of correct explanations produced between NP and ADJP idioms,  $t = -1.47539$ ,  $MSE = 3.13$ ,  $p = .1416$ . Thus, these results indicate that there are differences in the developmental course of idiom comprehension between VP and NP idioms, and between VP and ADJP idioms, but not between NP and ADJP idioms, when students are required to explain idiomatic meanings.

Further inspection of the results displayed in Figure 5 shows that there is a strong and consistent increase in students' explanation of VP idioms between both G2-G5 and G5-G8. For NP and ADJP idioms, on the other hand, the increase in accuracy of explanations appears strongest between G2-G5, with the rate of increase then tapering off between G5-G8.

To examine between which grades the growth of comprehension is significantly different for each idiom form, an additional set of post hoc analyses was performed. Results of the post hoc analyses of the between-grade differences, using Scheffe's procedures, are reported in Table 17. These results demonstrate that, for both VP and NP idioms, increases in comprehension continue steadily across the tested age range (for VP, significant differences between both G2-G8 and G5-G8; for NP, significant differences between both G2-G5, and G2-G8). On the other hand, for ADJP idioms, after an initial increase

between G2-G5, further positive changes were quite modest between G5 and G8 (G5  $\bar{M}$  = 30.2% correct, G8  $\bar{M}$  = 34.4% correct), hence the only significant age-related changes were between G2-G8.

To briefly summarize the findings for Questions #3 and 4: There is no single, general effect for idiom form. Rather, the results indicate that an interaction effect again occurs, in this case between response task and idiom form (i.e., significant differences among forms in the explanation task, no significant differences among forms in the item-selection task). In the explanation task, there are developmental differences among the three idiom forms used in this investigation. The developmental course for both NP and ADJP idioms differs significantly from the developmental course for VP idioms, while no significant differences in the course of development were identified between NP and ADJP idioms. Students are most successful in explaining the figurative meanings of VP idioms at each grade, and their success in explaining VP idioms increases more rapidly relative to both NP and ADJP idioms between G2 - G5 and again between G5 - G8. On the other hand, students' interpretations of NP and ADJP idioms follows a course that is similar for both of these forms, and students' explanations of the figurative meanings of these idioms increases only modestly between G5 - G8.

Now that the developmental differences between idiom forms have been established we next examine whether there are any Form x Grade interactions.

Question 5. If the developmental course of idiom comprehension differs based on idiom form, does the effect of form vary based on grade?

Item-selection response task.

For the item-selection task, as reported for Question 3, no main effect for idiom form was obtained, and there were no two-way Form x Grade interactions,  $F(4, 66) = .82$ ,  $p = .517$ ,  $MSE = .30$  (ANOVA5).

Explanation task.

Outcomes of the 3 (Form) x 3 (Grade) ANOVA6 reported in Question 3 did not identify any significant two-way interactions among Form x Grade,  $F(4, 66) = .86$ ,  $p = .492$ ,  $MSE = .77$ , indicating that the effect of idiom form on comprehension does not vary significantly by grade in this response condition.

The results presented to this point indicate that, depending upon the response condition, there are significant main effects for both analyzability (in the item-selection condition, and for the G5 and G8 participants in the explanation task) and for idiom form (in the explanation task). Next, Question 6 addresses the issue of whether there are any interaction effects among analyzability and idiom form.

Question 6. Are there interactions between analyzability and idiom form in children's acquisition of figurative meanings?

For Questions 6 and 7, the dependent variables are analyzability (SA - SU) and idiom form (VP, NP, & ADJP), and the independent variable is grade (G2, G5, & G8). A multivariate analysis of variance (MANOVA1) was conducted to test for possible Analyzability x Form interactions, and for any possible three-way interactions among analyzability, form and grade.

Item-selection response task.

Mean percentages of correct responses for analyzable and unanalyzable, VP, NP and ADJP idioms at each grade were calculated (denominator = 48 for each cell; range in the numerator = 29 - 47). The results are presented in Table 18, and displayed graphically in Figure 6. Visual inspection of the results in Table 18 suggests a complex pattern of acquisition, with comprehension differing on both of the within-subjects factors of analyzability and idiom form.

To test for statistically significant differences within this complex-appearing mosaic, a 2 (Analyzability) x 3 (Form) x 3 (Grade) MANOVA, with repeated measures on the first and second factors, was performed. The MANOVA yielded a significant Analyzability x Form interaction,  $F(2, 66) = 5.06$ ,  $p = .009$ , indicating that the effect of analyzability varies among the idiom forms.

To further examine this significant Analyzability x Form interaction, two sets of post hoc analyses were performed. First, Scheffe's procedure was used to identify between which grades significant increases in correct responses occurred for each of the 6 cells (SA-SU x VP, NP, ADJP). Then a series of dependent  $t$ -tests were performed to examine for the differential impact of analyzability on each idiom form. The Bonferroni procedure was again used to

correct for the effects of multiple  $t$ -tests (Harris, 1975).

Table 18

Mean Percentage of Correct Choices for SA and SU Idioms, by Idiom Form and Grade.

	<u>Idiom Form</u>							
	<u>Verb Phrase</u>		<u>Noun Phrase</u>		<u>ADJ. Phrase</u>		<u>Mean</u>	
	<u>SA</u>	<u>SU</u>	<u>SA</u>	<u>SU</u>	<u>SA</u>	<u>SU</u>	<u>SA</u>	<u>SU</u>
Grade								
2	62.5*	72.9	60.4	81.3*	60.4*	75.0*	61.1	76.4
5	72.9	79.2	62.5	91.7	70.8	89.6	68.8	86.8
8	91.7*	89.6	72.9	97.9*	85.4*	97.9*	83.3	95.1
Mean	75.7	80.6	65.3**	90.3**	72.2**	87.5**	71.1	86.1

Note. SA = Semantically Analyzable; SU - Semantically Unanalyzable

\* $p \leq .05$ ; \*\*  $p \leq .000$

The results of the first post hoc analysis, using Scheffe's procedures, are reported in Table 18. Significant increases in scores for unanalyzed NP and ADJP idioms between G2 and G8, and for analyzable VP and ADJP idioms between G2 and G8 were obtained. No significant between-grades differences were found for either unanalyzed VP or analyzed NP idioms.

Results from the second post hoc analysis, using dependent measures  $t$ -tests, showed that there were significant differences between SA and SU NP idioms, and between SA and SU ADJP idioms, with more correct choices for SU idioms for both forms. In contrast, the differences between SA and SU VP idioms were not significant,  $F(1, 35)$ ,  $MSE = .68$ ,  $p = .314$ . . Again, following from related comments in Question 1, while the factor analyzability interacted

significantly with the forms NP and ADJP, the functional effect was that unanalyzable idioms were identified more successfully than analyzable idioms.

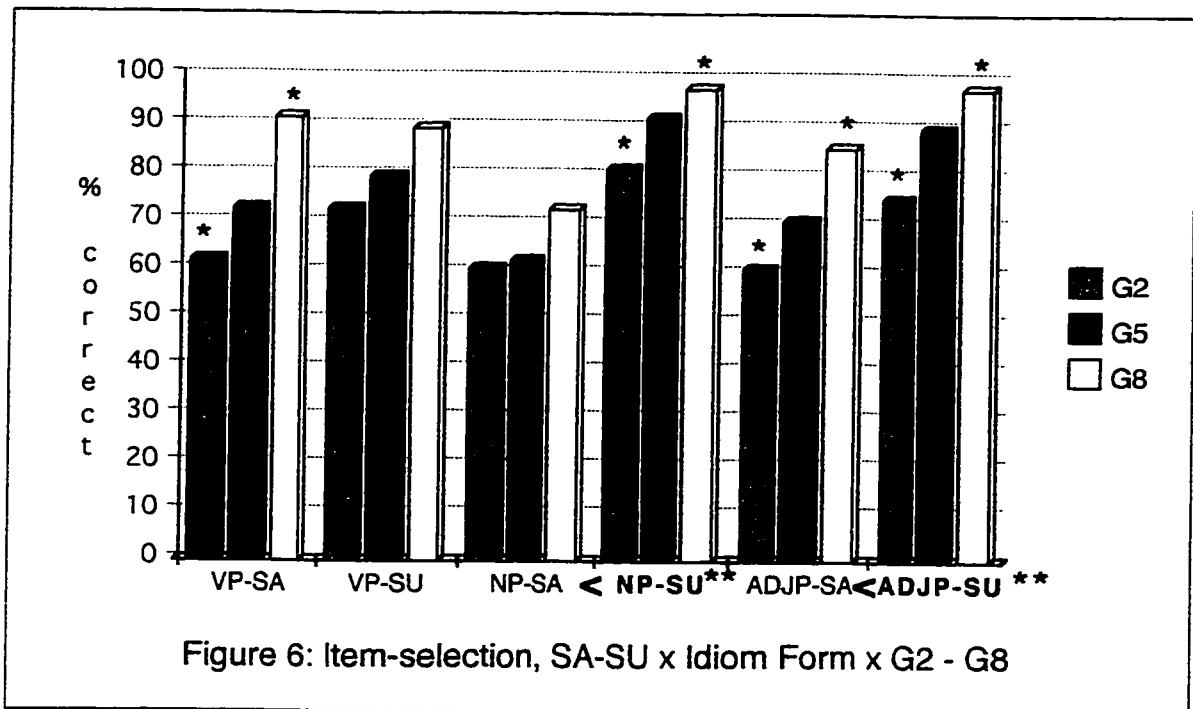


Figure 6: Item-selection, SA-SU x Idiom Form x G2 - G8

\*  $p \leq .05$ ; \*\*  $p \leq .000$  ( see Table 18 for specific values)

#### Explanation task.

Mean percentages of correct explanations for analyzable and unanalyzable, VP, NP and ADJP idioms at each grade were calculated (denominator = 48 for each cell; range in the numerator = 2 - 29). The results are summarized in Table 19. Visual inspection of the results again suggests a complex pattern of acquisition, with comprehension differing on both dependent variables. To test for statistically significant differences within this apparently complex mosaic, a 2 (Analyzability) x 3 (Form) x 3 (Grade) MANOVA(2), with repeated measures on the first and second factors, was performed. MANOVA2 did not yield a significant Analyzability x Form interaction effect,  $F(2, 66) = .35$ ,  $MSE = .28$ ,  $p = .706$ .

In Question 1, it was argued that the G2 students found the metalinguistic demands of the explanation task rather overwhelming, and an a posteriori ANOVA(4) was performed with the G2 students' explanation data removed.

Following from the reasoning presented in Questions 1, an additional a posteriori 2 (analyzability) x 3 (form) x 2 (grade) MANOVA(3) was conducted here to identify whether any Analyzability x Form interactions could be identified from the data for the G5 and G8 students. The results of MANOVA3 did not identify any significant Analyzability x Form interactions,  $F(2, 44) = .29, p = .747$ .

Table 19

Mean Percentage of Correct Explanations for SA and SU Idioms, by Idiom Form and Grade.

Grade	<u>Idiom Form</u>							
	<u>Verb Phrase</u>		<u>Noun Phrase</u>		<u>ADJ. Phrase</u>		<u>Mean</u>	
	<u>Analyzability: SA</u>	<u>SU</u>	<u>SA</u>	<u>SU</u>	<u>SA</u>	<u>SU</u>	<u>SA</u>	<u>SU</u>
2	16.7	16.7	10.4	04.2	10.4	18.8	12.5	13.2
5	39.6	31.3	25.0	20.8	35.4	25.0	33.3	25.7
8	60.4	45.8	35.4	31.3	35.4	33.3	43.8	36.8
Mean	38.9	31.3	23.6	18.8	27.1	25.7	29.9	25.2

Note. SA = Semantically Analyzable; SU - Semantically Unanalyzable

The outcomes of MANOVA1-3, conducted to identify interactions among analyzability and idiom form, indicated that once again, response type still matters in identifying whether (and where) significant differences in mean scores occur. In the item-selection condition, significant interactions among analyzability and idiom form occurred. Post hoc testing determined that the figurative meanings of unanalyzable NP and ADJP idioms were actually easier to identify than the meanings of analyzable NP and ADJP idioms. While students in G2 and G5 were somewhat more likely to correctly identify the figurative meanings of SU-VP idioms, this bias did not reach significance. No significant interaction effects were found among analyzability and form in the

results from the explanation task.

Several significant two-way interactions were identified in Question 6. The MANOVAs conducted in Question 6 also evaluated for any three-way interactions among Analyzability x Form x Grade, and these outcomes are reported next.

Question 7. Are there any three-way interaction effects among analyzability, idiom form, and grade in children's acquisition of idiomatic meanings?

Item-selection response task.

There were no significant three way interactions among Analyzability x Form x Grade,  $F(4,66) = .32$ ,  $p = .863$ ,  $MSE = .19$  (MANOVA1). Thus, the two-way Analyzability x Form interactions identified in Question 6 do not vary based upon grade in the item-selection response condition.

Explanation task.

As reported in Question 6 (MANOVA2, 3), there were no significant two-way Analyzability x Form interactions. No significant three-way interactions among analyzability x form x grade were identified,  $F(4, 66) = .47$ ,  $p = .757$ ,  $MSE = .38$ .

In the next section of the Results, the developmental course for ANOM idioms is outlined. Anomalous idioms are the *odds and ends* of the idiom closet: They are syntactically and/or semantically ill-formed, hence they don't fit well into taxonomic systems that attempt to categorize idioms based on either semantic (analyzability) or syntactic (VP, NP, ADJP, adverbials, etc.) metrics. Yet, these idioms are obviously both durable and frequently occurring in the language. As examples of durability, one of the ANOM idioms used in this study, *hither and yon*, has been in the language since A.D. 725<sup>70</sup>, another since the 14th century (*which is which* – O.E.D.), and several since the 16th century (*catch as catch can*; *tit for tat*). Obviously, at least the anomalous idioms studied here are not liable to *fall by the wayside*, and a linguist would certainly be remiss to term them *fair weather friends*.

Further, as was discussed in 'Methods', many ANOM idioms are in

<sup>70</sup> Oxford English Dictionary, 2nd. Ed.

common use. Indeed, they are common enough that from the corpus selected for familiarity ratings (and later augmented with additional ANOM idioms), it was not possible to match for familiarity with the VP, VP and ADJP idioms. Indeed for many of the ANOM idioms, familiarity ratings between 3 and 4 (somewhat common-to-very common) were the rule, not the exception. With these caveats regarding both their ill-formedness and their familiarity in mind, the findings for these idioms are presented next.

Question 8. How does children's acquisition of ill-formed or anomalous (ANOM) idioms compare with their acquisition of well-formed idioms?

For Question #8, descriptive data are provided to outline the development of figurative competence in anomalous idioms.

Item-selection and explanation tasks.

Mean percentage of correct explanations for anomalous idioms at each grade were calculated (denominator = 96 for each cell; range in the numerator = 26 - 81), and the results summarized in Table 20.

Table 20

Mean Percentage of Correct Item-Selection and Explanation Task Responses, Anomalous Idioms, G2 - G8.

Grade	<u>Response Task</u>						
	<u>Item-selection</u>			<u>Explanation</u>			<u>Mean</u>
	<u>M</u>	<u>SD</u>	<u>% correct</u>	<u>M</u>	<u>SD</u>	<u>% correct</u>	<u>% correct</u>
2	5.75	.97	71.9	2.17	1.27	27.1	49.5
5	6.5	1.31	81.3	3.67	1.23	45.8	63.6
8	6.83	.83	85.4	4.25	1.06	53.1	69.3
Mean	6.36	1.13	79.5	3.36	1.46	42.0	60.8

Note. Maximum score = 8 per task.

Visual inspection of Table 20 indicates that two of the broad trends identified in earlier sections of the results – (i) increasing accuracy in identification of idiomatic meanings with each increase in grade; and (ii) more correct responses in the item-selection task than in the explanation task – continue to hold true for the data from ANOM idioms. These idioms may be anomalous, but they behave like their ‘better formed’ relatives in at least these important ways.

Comparing the percentage of correct data presented in Tables 10 (SA - SU) and 20 (ANOM idioms), shows that students’ demonstrated very similar levels of overall accuracy in choosing the idiomatic paraphrase (ANOM  $\underline{M}$  = 79.5% correct; SA + SU  $\underline{M}$  = 78.6% correct). Additionally, students’ success at identifying the correct idiomatic paraphrases was quite similar at each grade (G2: ANOM  $\underline{M}$  = 71.9% correct; SA + SU = 68.8% correct; G8: ANOM  $\underline{M}$  = 85.4%; SA + SU  $\underline{M}$  = 89.2% correct).

Probably the most notable difference in the data between the ANOM and SA - SU idioms is in students’ explanations of figurative meanings in both G2 and G5. A notably higher rate of accuracy was obtained for the ANOM idioms (G2  $\underline{M}$  = 27.1% correct; G5  $\underline{M}$  = 45.8% correct) than for the SA + SU idioms (G2  $\underline{M}$  = 12.9% correct; G5  $\underline{M}$  = 29.2% correct). However, by G8, the advantage for ANOM over SA - SU idioms was, while still intact, more modest (G8 ANOM = 53.1% correct; G8 SA + SU = 40.3%). It is likely that the early bias in favor of the ANOM idioms was due in part to their greater familiarity, since, as the review of the literature has already indicated, familiar idioms are understood more readily than unfamiliar idioms (e.g., Nippold & Rudzinski, 1993; Nippold & Taylor, 1995). Why the early and strong advantage conferred by familiarity seemed to be eroding by G8 is, unfortunately, not immediately clear.

### Response task Interactions

The reasoning used in the Discussion to interpret the experimental results presented above led to the conclusion that an additional effect, an unexpected interaction between the dependent variables and the response tasks, had occurred. To test this proposed interaction statistically, the two response tasks were subsequently treated as a dependent variable (task), and

an additional MANOVA (4) was completed. MANOVA4 included a total of three within-subjects variables (analyzability, form, and task) and one between-subjects variable (grade). All the main effects and interactions obtained from MANOVA4 are presented in Table 21, along with estimates of power and effect size, to more completely characterize the interaction effects proposed in the Discussion.

Table 21

Multivariate Analysis of Variance for Interactions between Analyzability (SA - SU), Response Task (I/S - Exp), Idiom Form (Form) and Grade (MANOVA4).

Source	df	F	Power (@ .05 level)	$\eta^2$
<u>Between subjects</u>				
Grade (G)	2	20.72****	1.0	.557
<u>S</u> within-group error	33	(1.60)		
<u>Within subjects</u>				
Analyzability (SA)	1	10.29***	.874	.238
Form (F)	2	4.73*	.772	.125
Task (T)	1	492.21****	1.0	.937
G x SA	2	.94	.194	.054
G x F	4	.95	.284	.054
G x T	2	1.06	.219	.06
SA x F	2	2.60	.5	.073
G x SA x F	4	.56	.179	.033
<b>SA x T</b>	1	32.92****	1.0	.499
G x SA x T	2	.80	.175	.046
<b>F x T</b>	2	4.95**	.792	.13
G x F x T	4	.76	.231	.044
SA x F x T	2	1.92	.384	.055

Table 21 (continued)

<b>G x SA x F x T</b>	<b>4</b>	<b>.26</b>	<b>.104</b>	<b>.015</b>
SA x <u>S</u> within-group error	33	(.48)		
F x <u>S</u> within-group error	66	(.65)		
T x <u>S</u> within-group error	33	(.92)		
SA x F x <u>S</u> within-group error	66	(.82)		
SA x T x <u>S</u> within-group error	33	(.6)		
F x T x <u>S</u> within-group error	66	(.61)		
SA x F x T x <u>S</u> within-group error	66	(.56)		

---

**Note.** Values in parentheses represent mean square errors. S = subjects. Bold-faced entries refer to findings that are particularly relevant to arguments in the Discussion, following.

\* $p = .012$ ; \*\* $p = .01$ ; \*\*\* $p = .003$ ; \*\*\*\* $p < .000$

---

The statistical analysis of the overall results, as presented here in MANOVA4, was driven by the integration of the preplanned measures as discussed in the following section. Thus, they serve as a summary for the Results, and an introduction to the Discussion, to which we now turn.

## CHAPTER FOUR

### Discussion

The research contained in this dissertation was driven by three objectives. The first objective was to reexamine the impact of semantic analyzability on children's comprehension of idioms, by using only unfamiliar idioms to probe the development of figurative comprehension. Because the debate over the impact of analyzability on comprehension is unresolved (Table 2) it was reasoned that controlling for the effect of a key confound – familiarity – could add useful new data to this debate. The second objective was to identify whether, and to what degree, the developmental data on comprehension of VP idioms generalizes to NP, ADJP, and ANOM idioms. The goal of this portion of the investigation was to determine whether these other idiom forms behave in the same manner developmentally as VP idioms, the idiom form used almost exclusively in the research to date. The third objective was to determine whether the effect of analyzability varied based on grade, idiom form, or on both grade and form. The goal here was to identify and characterize all the meaningful stimulus, task and grade effects and interactions indicated by the data. Given the rather diverse developmental findings presented in 'Studies of Typical Development', above (see Figures 1-3; Table 2, Table 3), the use of such a multi-factorial approach appeared sensible.

The results from a series of preplanned and *after the fact* analyses on the data obtained from students in G2-G8 showed a complex pattern of effects and interactions among the dependent variables of analyzability and idiom form, and between the dependent variables and the task requirements of the two response conditions (item-selection and explanation). To evaluate, interpret and summarize these results requires a more detailed discussion of all the outcomes, and to this discussion we now turn, after first presenting a brief overall summary of the main findings.

#### Summary of the Findings

The main findings of this research can be summarized into three points:  
(1) the semantic analyzability of idioms does impact the comprehension of

idiomatic meanings; but analyzability interacts with response mode in determining whether analyzability facilitates comprehension (G5 and G8 students' were more accurate in explaining analyzable idioms' figurative meanings) or hinders comprehension (students across G2-G8 were substantially more accurate in selecting the figurative meanings for unanalyzable idioms); (2) the syntactic form of an idiom can impact the comprehension of idiomatic meanings, with the meanings of VP idioms identified most accurately; but this only occurred when students have to explain the figurative meanings; and, (3) analyzability interacts with idiom form; but only in an item-selection task.

In addition to these main findings, the present results are consistent with two of the general trends in the developmental research on understanding idioms: (i) comprehension increases systematically with increasing age; and, (ii) children demonstrate comprehension of idiomatic meanings more readily in item-selection tasks than when they are required to construct their own explanations for each figurative meaning (Ackerman, 1982; Ezell & Goldstein, 1991; Gibbs, 1987, 1991; Levorato & Cacciari, 1995; Prinz, 1983). An overall summary of significant effects is presented in Table 22.

Finally, the results of the prospective study of developmental patterns in anomalous idioms indicates that these idiom forms do not present unique comprehension challenges to G2-G8 students in either the selection or explanation task. Thus, while anomalous idioms are indeed structurally odd, they do not appear to be treated as 'anomalies' by the students in either comprehension task.

A detailed discussion of the outcomes of this research is presented next.

### Semantic analyzability impacts idiom comprehension -- but the nature of the impact varies with the response task

Perhaps the most unexpected and interesting finding of this investigation was the task-sensitive nature of the impact of semantic analyzability on idiom comprehension. The results show that analyzability does impact students' comprehension of idiomatic meanings -- but that analyzability also interacts with response mode in determining the specific nature of that impact. In some

circumstances analyzability facilitates comprehension- G5 and G8 students were more accurate in explaining analyzable idioms' figurative meanings. In another circumstance analyzability hinders comprehension of figurative meanings: students across G2-G8 were significantly more accurate in selecting the figurative meanings for unanalyzable idioms. Finally, analyzability may also have no effect on comprehension -- there were no differences between G2 students' accuracy of explanations of SA and SU idioms.

Table 22

Significant Main Effects for Analyzability, Form, Grade, and Their Interactions, by Response Condition.

<u>Factor</u>	<u>Response Task</u>	
	<u>Item-Selection</u>	<u>Explanation</u>
	<u>Significant @ <math>p &lt; .05</math>?</u>	
Analyzability	yes SU > SA	yes SA > SU, @ G5 & G8
Idiom Form	no	yes VP-NP; VP-ADJP
Analyzability x Idiom Form	yes SU > SA, NP SU > SA, ADJP	no
Grade	yes	yes
Analyzability x Form x Grade	no	no

While a number of studies have already identified impacts of response task requirements on when, and to what degree students demonstrate figurative knowledge, no developmental study has identified this significant interaction between response type and analyzability. Certainly, the present results were not predictable based on the semantic decomposition hypothesis, which assigns a facilitative effect to analyzability on comprehension. The outcomes also run counter to the preponderance of the relevant evidence regarding analyzability in the developmental literature (Gibbs, 1991; Nippold & Rudzinski,

1993; Nippold & Taylor, 1995). How then, to account for and integrate these outcomes with the previous research on analyzability and the development of idiom comprehension, particularly given the seemingly contradictory results of the present research, as outlined in Table 22?

### The ecology of idiom comprehension

First, it is argued that these results need to be examined from within the framework of narrative (discourse) processing, rather than from within the framework of events occurring at the phrase or sentence level. While a phrase or sentence level analysis is both traditional and built into the design of research on idioms, the present results indicate that use of only this level of analysis is inadequate for explaining the results. A productive accounting of the current results requires that we consider and integrate multiple factors influencing children's comprehension of messages, factors associated with the fluid frame of ongoing discourse<sup>71</sup>. This requires that we acknowledge that children don't set out to analyze idioms, but rather to understand messages. In addition, how they seek to understand can also be influenced by what they have been directed to do in assessment tasks. Use of these perspectives then requires that we acknowledge that a semantic analysis is only one of several levels of analysis. Hence, any effect of analyzability has to be evaluated from within these multiple levels of analysis, and within the broader network of factors impacting idiom comprehension in larger units of text. These arguments are discussed below in greater detail, after a brief general overview of discourse processes judged to be relevant to participants' comprehension of idiomatic meanings.

The discourse factors tentatively identified as relevant in processing the unfamiliar idioms used here include the following:

i) factors typically associated with dyadic communication. In particular, the perceived (or received) purpose/s of the communication, the roles and relationships of the investigator and student, and explicit or implicit assumptions each student makes about what is required of his/her comprehension and

<sup>71</sup> However, the need for a multidimensional analysis of idiomatic meaning and use is certainly not novel. For a thorough discussion of the syntax, semantics and pragmatics of the conjunctive idiom *let alone*, see Fillmore, Kay, and O'Connor (1988).

production in a particular exchange. These factors are sociolinguistic elements of the communication;

ii) text-level information, provided by the narrative that each idiom is embedded within. This includes information about the setting, character(s), problem and resolution, particularly as the commentary or resolution relates to the idiom within the narrative; and,

iii) phrase-level information related to the target idiom – the degree of analyzability and familiarity, and possibly the syntactic form – that is, the psycholinguistics of the idiom.

With the above as a framework, and remembering that the participants are also reacting to the specific demands of each comprehension task, let us now use this framework to examine the current results.

#### The social context.

First, consider the social context of the data collection. Participants were seen on two separate occasions. In the first meeting, the investigator developed a scholastic testing environment. The adult was 'in charge' of the interaction in administering a series of standardized test measures. The pace, content and flow of information were tightly controlled by the investigator. This interaction can be (crudely) scripted as: 'you listen – I say – you figure (something) out – you answer'. No negotiated or shared meaning was possible. Each participant was, therefore, required to rely on his/her own skills and knowledge, and on whatever information was provided within each story to develop the response required by a particular assessment task<sup>72</sup>.

In the second session this basic dyadic pattern was maintained. The investigator read the narratives with the target idioms, the participants listened, processed a modest amount of verbal information, and came up with a required response. Students were told that the investigator would not provide any additional information. Thus, by the time the students first came in contact with a target idiom, it is postulated that they at least tacitly understood they were 'on their own' in terms of the resources available for answering each question. In both idiom comprehension tasks this meant attending to, and extracting

<sup>72</sup> As outlined in Methods, several of the subtests had processing demands that were deliberate analogs to those encountered during data collection – Verbal Comprehension and Paragraph Comprehension in particular.

information from (1) the story context, and (2) the idiom within the terminal sentence. Next, we consider the text-level information that was available, and common to both response tasks.

The narrative context: text-level information.

The story context for each idiom consisted of two or three sentences plus a terminal sentence containing the target idiom as the sentence-final phrase. Each story contained a character(s), setting and problem, combined to describe a short behavioral episode that lead plausibly into the sentence containing the target idiom, and supporting an idiomatic interpretation of the target phrase. For example, the story line for the idiom *asleep at the switch* (SA-ADJP; Appendix C) runs: "The game was tied, but the bases were loaded. Unexpectedly, Kevin watched as strike three came in belt high right across the plate. Later the manager said that Kevin was *asleep at the switch*."

Now, the story does not directly produce the idiom's paraphrase of 'failing to act promptly, as expected' (the context also supports paraphrasing the idiom as 'he lost the game', 'he choked', and so forth). However, the text does support elements of the figurative meaning- that is, an important moment, a missed opportunity, and the implied assignment of failure due to an individual's inaction. Thus, a listener can use this text-level information to aid in identifying in general the figurative sense of the idiom. Certainly, if the idioms used are truly as unfamiliar as has been argued in Methods, then the contextual information must provide substantial information toward solving the idiomatic puzzle. Arguably, both the total amount of, and balance among, students' total figurative explanations (that is, C-I + F-O; Table 14: MSA = 68.3%; M SU = 64.1% correct) speaks toward the participants' use of context, and of the equivalence of the contextual information provided for both the SA and SU idioms. This assumption regarding the informativeness of the context is consistent with much of the research done with children in this age range (Figure 3), and argued with particular force by Levorato and Cacciari (1992).

While the information provided by context may be a necessary component in describing the students' understanding of idioms, it is not sufficient in differentiating among the current findings. To do this, we must also include the information contributed by each idiom, stipulating that analyzable

idioms provide more semantic information than unanalyzable idioms. Next, we evaluate how the information provided by each idiom interacts with the task demands and with students' 'test-taking' responses to each comprehension task. It is to this that we turn next.

Analyzability— the contribution of the idiom as a complex unit.

First, consider that analyzability is, as the term implies, a dynamic characteristic of a complex unit such as an idiom. To say that an idiom 'is analyzable' is to say that an idiom's meaning is distributed across separate elements of meaning that independently contribute to the overall figurative sense. However, these separate semantics elements are not 'spring loaded' -- they aren't available unbidden. Instead, idioms are 'analyzable' only to the degree that the listener uses some sort of processing strategy that uses those elements of meaning while attempting to (re)construct the idiom's figurative meaning *in light of* some communicative or task demand.

With the above in mind, also recall that the target idioms were heard as the final phrase in the story, and the students then had to immediately explain in their own words the meaning of each idiom. So, the communicative demand on the subjects was to provide an accurate and independent rendering of the meaning of unfamiliar phrases encountered within each story, and to do this in their own words. The task demand in the explanation condition was two-fold: first to identify or decode the figurative meaning, then use their own semantic/lexical resources to develop or encode a (literal) phrase or sentence whose meaning is synonymous with the figurative meaning of the idiom.

Paraphrasing figures of speech: The impact of context, text and analyzability on students' explanations of idiomatic meanings.

When the G5 and G8 students assembled their own (literal) explanations for each idiom's figurative meaning, this effort was more successful with analyzable idioms. This facilitative effect occurred because analyzable idioms' "individual components share the same semantic fields with their idiomatic references" (Gibbs, Nayak & Cutting, 1989, p.578). It is this conceptual congruence, akin to synonymy, which facilitates the more successful recruitment of related terms and meanings that students can use when assembling their

explanations.

To illustrate, let us again consider the example of *asleep at the switch*. The text provides some information, as detailed above, and some of the students indeed focused their analysis at this level, as their explanations of the final phrase reflect:

“like he's lousy or something”;

“that he was just going out of his way, that he wasn't performing expectedly”

These explanations again reflect that the text supports assigning failure to an individual.

Now, while the information provided by the text is useful, it obviously does not over-determine the correct figurative explanation. Children will (and certainly did!) propose their own solutions about what the meaning of these odd phrases might be. For example, in the next explanation (of *carry coals to Newcastle*) the G2 student relied heavily on a combination of text and imagination, providing a reaction to, or resolution of the story without ever directly defining the idiom:

“oops, I already bought, she already bought two cases of cans, and I told mom nobody bought two cases of cans, and mom was furious, because that was the money we were using to go to Disneyland”

Clearly, this student brought some of her own resources to bear in working out that the idiom refers to doing something unnecessary!

As argued earlier, analyzable idioms do contain additional semantic information that can be used to go beyond the more general text-level information. Using this data, the students can and do work to identify and explain the figurative meanings. This focus on the elements comprising the idiom can be seen in explanations provided by other G2 students trying to ‘work out’ the meaning of *asleep at the switch*:

“he falls asleep like a light bulb”;

“very fast; fall asleep very fast”

Here, we can see that ‘switch’ apparently contributes the sense that the event is sudden, and that something about a desired action must occur rapidly (but it's not always clear of just *which* action...). Again, of course, as was the

case with the text-level information, the information contributed by the semantic components of analyzable idioms may mislead, as well as lead. This can be seen in the explanations offered by yet other students (in the spirit of working out an unfamiliar phrase, the idioms associated with each explanation are footnoted):

“a stinger; lie and fib; that’s why I said a stinger; a fib is called a stinger”<sup>73</sup>

“slept with the chickens; no, uhm, (that’s a joke); went to bed with, uhm, out saying a word”<sup>74</sup>

As the guys at the carnivals say, ‘close, but no cigar’.

And sometimes, the additional semantic information provided by the idiom can lead a student on a rather fanciful journey, but one that ends in *more or less* the appropriate place:

“she went to bed to chickens means, means, she probably lives on a farm and, she looked out for the chickens, and when the chickens went to bed she raced and got her pajamas on and went to bed so she could get up early”

The overlap of semantic/conceptual meanings between the elements of analyzable idioms and their figurative sense, which facilitates students’ explanation of idiomatic meanings, would not occur in unanalyzable idioms. For SU idioms, there is a completely arbitrary relationship between the words comprising the (unanalyzable) idiom and the idiom’s underlying figurative meaning. This is because unanalyzable idioms have no internal semantic structure, hence no semantically independent elements. In this study, for example, the individual referents of the words comprising *ring the changes* (SU-VP), the *life of Riley* (SU - NP) and *all wool and a yard wide* (SU - ADJP) have no motivated relationship with ‘repeat the same idea in many ways’, ‘a soft, easy life’, and ‘of fine character; especially, very generous and kind-hearted’ respectively. In this circumstance, any attempt to do a compositional analysis would, by definition, fail.

On encountering an unanalyzable idiom, the listener is forced to either process the idiom as a single, unanalyzed semantic unit, or to at least attempt a compositional analysis. Any compositional-first processing would fail (Abkarian,

<sup>73</sup> She has a bee in her bonnet - SA-NP.

<sup>74</sup> go to bed with the chickens SA - VP

et al., 1992), and in addition would likely consume additional processing time, as Gibbs, Nayak, and Cutting (1989) found in a reaction time study with adults. Thus, the respondent has to explain the SU idiom's meaning using only the text-level information. With unfamiliar SU idioms therefore, the student is at a disadvantage in terms of the total information available relative to the total information available for SA idioms. Hence the advantage for SA over SU idioms in the explanation response condition, for G5 and G8 students.

Notably, the explanation offered above as accounting for the SA>SU significant difference does not extend, either in this argument or in the data, to the explanations offered by the G2 students. Reprising an argument first offered in the Results, this is because the metalinguistic demands of the explanation task itself were so challenging that these young students had significant difficulties in encoding an appropriately accurate response. The difficulty may lie primarily in the encoding of, not the decoding of, the figurative meaning. This is so, because the responses in item-selection, a 'match the meanings' decoding task, were rather accurate ( $M = 68.8\%$  correct), while their explanations obviously were not, at  $M = 12.9\%$  correct.

This finding and conclusion for the G2 students is consistent with the related developmental evidence regarding the challenges of explanation tasks shown in Figure 2. In addition, examination of a simpler, but analogous explanation task, the Oral Vocabulary subtest from the TOLD-2 Primary (Newcomer & Hammill, 1988), shows that explanation is indeed a difficult metalinguistic task for students between 7;0-7;11 years of age. In the grade two age range, obtaining a score within normal limits (25th percentile) only requires a student to correctly define 11 relatively straightforward items. These include such familiar everyday terms as 'bird, apple, brown, ice, cow, bed, zoo' and so forth, terms whose referents are presumably much more familiar and available to these students than *bred in the bone*. Thus, it is not really surprising that these seven year-olds did not successfully make use of the conceptual underpinnings of analyzable idioms. They had plenty on their hands just figuring out *which* end was *which*.

Matching meanings: The effects of context, text, and task demands in item-selection.

Now, let us return to the discussion of the results from the item-selection task. In this task, and in contrast with the explanation task, SU idioms were identified more successfully than SA idioms. It is argued that this is due to: (1) when the response task occurred, (2) the nature of the task, and (3) conflicts between the likely level of analysis and what would be maximally effective response strategies.

To recall: The item-selection task was administered for each idiom immediately after participants had provided their own explanation of the idiom's meaning. After providing this explanation, participants heard the target idiom again in a carrier phrase ('does \_\_\_\_\_ mean...'), then listened to a list of three alternative paraphrases, and were directed to select the paraphrase with the same meaning as the target idiom.

The communicative demand on the students was to listen carefully to the three alternative paraphrases, then to tell the investigator which one choice went with the idiomatic phrase. The task demand in the item-selection condition was three-fold. First, hold the target idiom in working memory, then compare the student's own gloss of the idiom's meaning (presumably derived during the story and explanation sequence) with each alternate definition, and finally to select the 'best' match. Encoding requirements were modest, requiring only a reporting of choice 1-3 to the examiner.

Under these circumstances, a substantial load is placed on working memory in the item-selection task. The student has to hold the target idiom in memory, rapidly compare the idiom's meaning with the meanings of each of the three paraphrases in turn, and hold conditional decision data at hand until the final selection is made. In this circumstance, a 'solution' to the challenge of selecting the correct paraphrase would be to identify the item that either (1) matches the meaning generated by the students during the explanation task, or (2) which sensibly completes the narrative. It is argued that, when the figurative meaning is unknown, there is an advantage in treating these phrases as unitary chunks, instead of attempting to further analyze their meanings. By matching each idiom as a 'whole' unit of meaning with each alternative paraphrase, the

selection process, which is working to identify the paraphrase that works in re/solving the narrative, is likely to function with maximum speed and accuracy.

A visual analogy to this task is found in an almost completed puzzle. When there are only 3 or 4 pieces left, generally the most efficient strategy is to try to match the pieces and 'holes' by outline and overall fit, not by connecting the individual lines on each piece with each possible extension on the board.

By definition, unanalyzable idioms are the best candidates to be treated as single, 'whole' units of meaning. Treating them in this fashion would, as argued above, maximize the 'hit rate' for accuracy, and the data do show that in item-selection, the meanings of SU idioms are identified more accurately than SA idioms. On the other hand, attempting a compositional analysis of the components of analyzable idioms' meanings in item-selection can be a strategic processing error. Attempting such an analysis imposes additional demands on immediately available cognitive resources. As outlined in the Review of the Literature, a wide-ranging analysis can consume more processing time, as each component is mapped and the various related meanings are sorted through and integrated into the overall meaning of the text. Thus, choosing a semantically detailed analytic strategy may in effect induce "paralysis by analysis", whereby the item-selection (rapid) choice process is temporarily paralyzed by the analysis of the individual elements of meaning.

To recapitulate: In the explanation task, analytic processing is effective because students have to first decode, then reconstruct and explain the figurative meanings. The process of reconstruction-explanation benefits from the variety of semantic-conceptual data encountered during the analysis of an idiom's individual components. The advantage shown for SA idioms in the explanation task demonstrates that these idioms' do have individual semantic components that participate in comprehension processes.

On the other hand, in the item-selection task, unitary or gestalt processing is more effective. This is because the students are first looking to identify the paraphrase that works to complete the meaning of the narrative, since the unanalyzable idiom does not independently contribute any additional information. Attempting to analyze the elements of an unanalyzable idiom temporarily 'paralyzes' the more efficient gestalt matching strategy, thus

depressing the accuracy rates for SA idioms in item-selection. Again, if you stop to analyze the colors and lines in a piece of a nearly completed puzzle, instead of looking for the piece that fits, you do take longer to finish the puzzle! Finally, the advantage shown for SU idioms in the item-selection task indicates that these idioms do not have individual semantic components that can participate in the comprehension processes.

Together, these results support the thesis that young children do not set out to parse idiomatic meanings, but rather to make sense of discourse, and to respond sensibly to the demands of specific communication tasks.

The syntactic form of an idiom impacts the comprehension of idiomatic meanings, but this effect occurs only when students explain the figurative meanings

Another interesting outcome from this research was the finding that the syntactic form of an idiom can also impact idiom comprehension. The purpose of this portion of the experimental design was to explore whether VP idiom data can stand as a proxy for the behavior of all idioms forms. That is, is it reasonable to suppose that developmental data from the comprehension of VP idioms generalizes, a priori, to every other idiom form? The short answer, based on these data is no; the data from VP idioms does not adequately describe the developmental course of either NP or ADJP idioms.

The results show that the syntactic form of an idiom impacts comprehension -- but the effect for idiom form only occurs in the explanation task. The differences in comprehension between forms -- with students correctly identifying more figurative meanings of VP idioms than of either NP or ADJP idioms - are detected only in the explanation task. Conversely, in the item-selection task there is no impact of idiom form, as Figure 4 clearly shows.

This portion of the research was designed to explore whether there was in fact any impact of idiom form on comprehension, not to test a specific prediction. Thus, it is not immediately clear why differences in comprehension among the three idiom forms occurred, and occurred only in the explanation task. Nevertheless, possible explanations about why an effect for form was found are offered, while acknowledging that a more theoretically compelling

account awaits both replication and a more detailed examination of this particular finding.

When a significant effect for a factor (idiom form) is found in only one response condition (explanation), this again suggests that an interaction occurs between that factor and the response tasks. So, let's begin by exploring this possibility.

In the section of the discussion devoted to analyzability, a significant effect for the then factor of interest (analyzability) was found in item-selection (all grades), and in the explanation task (for G5 and G8). However, no effect for idiom form can be identified in item-selection. Indeed, the results show that the developmental progression for VP, NP and ADJP idioms is practically isomorphic. Why?

I believe that this is because, unlike the case with analyzability, there is no 'paralysis by analysis' occurring with any syntactic form. That is, for analyzability there are theoretically sensible reasons for arguing that the attempt to analyze the individual components of an unknown phrase's meaning might 'paralyze' the most efficient processing strategy for selecting a correct answer. However, there is no particular reason to believe that an analogous effect for lexical/semantic search would occur among the different syntactic idiom forms evaluated, and indeed no such effect occurred. In item-selection the students tackle each form in the same manner, and there is no effect.

Yet, this portion of the argument does not suggest why students were more successful in explaining the figurative meanings of VP idioms, when compared with both NP and ADJP idioms. Why a significant effect for an idiom's syntactic form in the explanation task? Why should it be easier to provide a literal paraphrase for the figurative meaning of a verb phrase idioms than for either a noun phrase or adjective phrase idiom?

The first possibility to consider is that it is not just individual idioms that vary in familiarity, but that whole categories of idiom forms also vary in familiarity<sup>75</sup>. Idioms occur more in some syntactic forms than in others, and this may mean that students are more familiar with some idiom forms per se, than with others. A high level of familiarity with a particular form may lead the listener

<sup>75</sup> and it is possible that this effect may be only, or most noticeable when unfamiliar idioms are used as stimuli.

to more readily identify that particular form as being nonliteral, and this may confer some (otherwise unspecified) processing advantage on the more familiar form. Thus, comprehension of VP idioms as a syntactic class of idioms may be advantaged because students encounter VP idioms more often than they encounter other idiom forms day-to-day. It is this familiarity with the form in general, not any idiom in specific (particularly with unfamiliar idioms), which would therefore provide the advantage for comprehension.

While it is not possible to prove this hypothesis here, there is at least modest evidence to support one portion of the argument: VP idioms are commonly occurring forms. For example, as mentioned previously, verb/verb phrase idioms are the most commonly occurring idiom forms in the idiom dictionaries reviewed for this study, providing approximately 50% of all entries (please remember that this is only an armchair estimate). Thus, VP idioms represent the predominant Type of idiom form, at least based on the dictionary data. And, many VP idioms have been rated and/or selected by adult raters as being very familiar in a number of prior adult and child idiom studies. This indicates that at least a moderate number of different VP idioms are frequently occurring Tokens. Finally, in a humorous and clever corpus of complex forms collected by Ray Jackendorf and his daughter Beth, the WoF corpus<sup>78</sup> (1995), some field data on the distribution of idiom types used in one public venue is available. A total of 66 idioms were listed in the WoF corpus. Of these, 42% were VP idioms, 20% NP idioms, 17% ANOM idioms, and the remainder a variety of other idiom forms, again indicating that VP idioms occur frequently as an idiom form.

The proposed effect for familiarity of form, conferring some advantage for VP idioms with students engaged in paraphrasing the figurative meaning, is highly speculative. The major advantage would seem to be to signal more quickly to students that the meaning of the target phrase should be considered as a function of the information available at the text-level, rather than first working to decompose the idiom. Whatever the impact, students seemed to capitalize on it across the grades. Re-examination of Table 17 shows that the advantage for explaining the meanings of VP idioms relative to both NP and

---

<sup>78</sup> for Wheel of Fortune, the source of the corpus.

ADJP idioms is found first in G2 (VP explanations had 14% higher accuracy), and that the absolute difference in percent correct explanations grows for VP idioms at each grade. By G8, VP idioms were explained accurately at rates more than 50% higher than the explanations for ADJP idioms, the form with the intermediate accuracy rate. Indeed, if VP idioms do occur much more frequently than other idiom forms, it may be that this advantage is cumulative, as students encounter more idioms in literature as well as in the language of the classroom, and the data reflect this cumulative impact.

Semantic analyzability interacts with idiom form, but this effect only occurs in the item-selection task

The particulars of this interaction shed some light onto that profound question, elucidated most clearly by those social linguists, Bud Abbott and Lou Costello, "who's on first?". Or, more prosaically, when both idiom form and analyzability interact with each other, and with the response tasks, can we identify whether, in the joint interactions between analyzability, idiom form and the response tasks, particular interaction effects have priority? A logical analysis of the pattern of results obtained, combined with the reasoning used to this point in the Discussion, indicates that a model for order of effects (or priority, or power) can be developed, and this is discussed next.

To review, the flow of effects and interactions can be mapped as:

1) analyzability interacts with response task-

SU>SA in item-selection;

SA>SU in explanation;

2) form interacts with response task-

VP  $\simeq$  NP  $\simeq$  ADJP in item-selection;

VP>NP and VP >ADJP in explanation;

Now, combining these results in the form of an equation, to help visualize the two-way interactions, we see that:

3) analyzability interacts with form -

Item-selection:

$$(SU>SA) \times (VP \simeq NP \simeq ADJP)^{77} = SU NP > SA NP$$

<sup>77</sup> the () are used to code each effect: (the effect of analyzability) x combined with (the effect for form) in each response task condition as indicated

SU ADJP > SA ADJP

but: SA VP  $\simeq$  SU VP (hence, the  
interaction effect)

Explanation:

(SA>SU) x (VP>NP, VP>ADJP) = no interaction effects

SA (all forms)>SU (all forms)

From this schematic, we can see that both the dimensions of analyzability and response task are powerful. And, as has already been argued, the response requirements of each task bring out different aspects of the impact of task on both analyzability and idiom form. Thus, we can conclude that the impact of the response task is very important. This conclusion is supported by the differences in effect size obtained in MANOVA4 (Table 21), the MANOVA completed after the fact to more precisely quantify the interaction effects under discussion here. Approximately 94% of the variability in scores can be attributed to the effect of the response task, while important but much less dramatic amounts of the variability in scores are accounted for by analyzability (24%) and idiom form (12.5%)<sup>78</sup>.

Next, when we consider the impact of the interaction of task and analyzability on idiom form in item-selection, we see that the unanalyzability of idioms (more correct choices in item-selection) specifically facilitates comprehension for two out of three forms (NP and ADJP idioms). Further, when analyzability meets form in the explanation task, the positive effect for analyzability does not vary significantly across the three idiom forms. In each of the three idiom forms, students are more likely to succeed in explaining the meanings of SA idioms than the meanings of SU idioms. As has also been argued previously, this is likely due to the general facilitative effect of analyzability when students have to locate and encode literal synonyms to explain the meanings of figurative phrases.

Therefore, based on the evidence above, and from the data on both power and effect size in Table 21, we can say that the effect of analyzability has, in some sense, priority over the effect of idiom form ( $\eta^2$  for analyzability = .238,

<sup>78</sup> While  $\eta^2$  is generally considered a biased measure of effect size (Howell, 1989), for our purposes in considering the relative magnitude of each variable and interaction the statistic is useful.

for form = .125). However, one question remains: In item-selection, why did the behavior of the VP idioms differ from that of NP and ADJP idioms? Specifically, why was there no significant difference in the likelihood of students' selecting the correct paraphrase between SA and SU VP idioms?

Again, no certain answer is available. One explanation suggested by the overall pattern of the results is that students are more likely to use a 'bottom up' or analytic strategy when trying to identify the meanings of both SA and SU VP idioms, even when this strategy is ultimately less productive in item-selection. It may be that, as students work to first explain meanings, the two-way interaction between text level and phrase level information has a longer time course when VP idioms are encountered than when either NP or ADJP idioms are encountered. This might occur because (a) the proposed familiarity for VP idioms as-a-class cues the need to scan both phrase and text level information for alternate (nonliteral) meanings, an extension of the argument proposed in the previous section on idiom form; or (b) because the syntactic structure of VP idioms promotes, in some sense, a broader and more general search for related meanings. In either (a) or (b), this behavior would likely be helpful for constructing explanations, but potentially obstructive when a top-down, 'whole - to - whole' matching strategy could place fewer demands on working memory.

But why would the syntactic structure of VP idioms 'promote' a more general search for meanings? Perhaps because the syntactic structure of VP idioms is different in an important way from either NP or ADJP idioms. Following from Everaert, VP idioms are "idiomatic phrases that consist of an argument-taking verb and its arguments, whereby the verb has at least one open argument position or one of its arguments contains an open argument position" (cf Van de Voort & Vonk, 1995, p.28). Here, the listener may perform additional linguistic-cognitive work in an attempt to fill the hypothesized empty argument position, with this being attempted for both SA and SU idioms. Again, this may 'pay benefits' when the communicative demand requires students to explain a figurative meaning using their own words, but extract a cost when the task requires the listener to rapidly compare the target with three alternative meanings, without recourse to any additional information. This structural condition for VP idioms does not generally hold for either NP or ADJP idioms,

and there would not be any additional work to 'fill' an open argument position for either of these forms.

The findings for main effects, for interaction effects, and of unexpected effects, all serve as an effective (but accidental) metaphor in highlighting the challenges the young, intuitive linguists in this study faced in 'figuring out figurative meanings'. The outcomes also serve to demonstrate the complexity and reactivity of many of the elements that contribute to understanding idioms. The powerful impact of the response tasks, and the surprisingly multidirectional interactions that occurred with between the response tasks, and both form and analyzability, all underline the need to investigate idiom comprehension using a multi-axial or multi-dimensional approach. Some of the less potent effects (of form, in item-selection, for example), still raise interesting questions, questions which await the replication of these results.

The complexity and reactivity of several of the elements that contribute to typically developing children's understanding of idioms have implications for the assessment and treatment of children with atypical language development. It is to this area that we turn next.

### Clinical Implications<sup>79</sup>

There is a good deal of clinical interest in evaluating and treating children who experience difficulties in developing idiomatic competence. While his study was not designed to address clinical issues, the data can be used to generate some suggestions about the evaluation and treatment of children with atypical figurative language development, and several suggestions will be presented. Clinical implications will be discussed in two parts: (1) implications for assessment, and (2) treatment implications.

#### Assessment

The explicit assessment of idiomatic competence is often not done until students are in the middle grades, generally between grade five and grade eight. The first goal in assessment is usually to determine if development is

<sup>79</sup> The discussion of clinical implications is generally focused on school-aged students who participate in regular education classes. Specifically excluded is consideration of the clinical needs of the Deaf, issues related to adult second-language learners and adult aphasics.

typical or atypical, with more detailed testing done only when needed to develop treatment objectives. Standardized assessment measures used by SLPs include: (1) the Fullerton Language Test for Adolescents (Thorun, 1986), which requires students to define the meanings of an assortment of idioms in a no-context condition; and (2) the Test of Word Knowledge (TWK) (Wiig & Secord, 1992). The TWK requires students to either select the figurative paraphrase for a given idiom, or, when given the figurative meaning, to select the correct idiom from among four idioms. These measures are used primarily to identify atypical development, and the degree of delay. Used in concert, they can provide more comprehensive information than either alone, but because they use different target idioms and have different normative bases, direct integration of the results is problematic.

The present results support including the following, in common for both the identification of a deficit and for assessment done to identify specific treatment objectives:

1. Assessment tasks should present idioms within short story contexts. Identifying idioms in isolation is obviously pragmatically unreal on the face of it, and our primary interest is in determining students' abilities to understand idioms within discourse. If there is an indication that a student has a unique difficulty in integrating narrative-level information, then separate assessment of idioms in isolation and in context may be required. Because knowledge of idioms is highly language- and cultural-specific, the use of idioms in isolation will tend to produce a higher false-positive rate, particularly given the diversity of many school populations. On the other hand, presentation within a story context may more accurately reflect students' real-world abilities to map new meanings within discourse.
2. Tasks should require students to both explain figurative meanings in their own words, and use an item-selection task within the same measure (the TWK's use of two selection measures -- 'here's the idiom, you find the meaning', and also 'here's the meaning, you find the idiom', is a very appealing clinical procedure!). The challenges of an explanation task can obviously result in underestimating many students' comprehension of figurative meanings, at least through grade eight, while item-selection may provide too generous an estimate

of figurative competence. Both tasks are needed to develop a more nuanced picture of competence than would be provided by either task alone.

3. Tasks should use a mix of analyzable and unanalyzable idioms, given the demonstrated interactions between analyzability and response task; and finally,

4. Tasks should use a mix of idiom forms that are more representative of the diversity of idioms. Certainly, these should include both VP and NP idioms, given the very different growth curves for these idiom forms.

### Treatment

There are few references in the literature addressing the treatment of idiom comprehension deficits (see Ezell & Goldstein, 1991, for apparently the only study containing treatment data). Indeed, the lack of data regarding when, why, how and to whom treatment is delivered, is rather comprehensive<sup>80</sup>. About this issue, it seems, we are almost completely *in the dark*, at least in terms of language pathology as a clinical science. Since the basis for clinical service is essentially undefined, the suggestions offered in this section are generic. They are offered as examples of where the present results seem to point in terms of treating students who are having difficulties in comprehending idioms (and likely other figurative forms), and who participate in regular education classes.

These results suggest that, for young English first-language speakers, clinicians might consider whether, upon finding 'deficits' in idiom comprehension, targeting idioms for direct instruction is an appropriate first-step. The results indicate that the comprehension of idioms is a complex, multidimensional process, and one embedded within the multi-level framework of discourse processes. What appear as obvious deficits – the inability to figure out idiomatic meanings – may actually be a symptom of larger difficulties in extracting meaning from text. On this basis, interventions might need to address both narrative skills (Ezelle, 1998) as well as provide regular contact and practice with identifying, explaining and using these commonly occurring figurative forms. Habilitation of idiom comprehension deficits would therefore occur as but one component of a sequence of treatments, arising within the broader framework of treating deficits in comprehending narratives.

<sup>80</sup> But see Nippold (1991), for some interesting and sensible suggestions regarding general approaches to intervention.

As students develop skills in narrative comprehension, I would suggest then introducing frequently occurring, unanalyzable VP idioms within story contexts at the first level of training specific to 'idioms'. Frequently occurring idioms would be used because one goal is to efficiently develop skills with a high impact on daily functioning. At this introductory level, helping the student identify that an idiom: (a) is a complex unit, and (b) has a meaning different from the words comprising the idiom itself, may best be achieved by using the unanalyzed form. As has been argued above, there can be a 'paralysis by analysis' in attempting a compositional processing strategy. Using unanalyzable idioms will force a student to attend to the discourse and text levels of information first; that is, to use the newly learned narrative skills to at least 'narrow the search' in identifying the figurative meaning. This could start with brain-storming possible meanings based on the narrative and text levels of information, then having the student select the appropriate figurative gloss from a list of possible choices.

When the student reaches some criterion of success indicating that s/he has reasonable facility in using discourse and text level information to help decode the idiom meanings, one might then introduce analyzable idioms of the same syntactic forms for the next level of training. However, to help the child to avoid being garden-pathed into relying on a compositional analysis, idioms that have plausible literal and figurative interpretations should be excluded at this stage. That is, don't use *kick the bucket*, *spill the beans*, and so forth, but rather analyzable idioms that have only plausible figurative interpretations. While we now want to train students to begin to pay attention to the internal semantics of familiar idioms by this stage, it is the smaller step in a training sequence to first use idioms that can only be glossed figuratively. We are introducing to the students an additional strategy -- the possibility that figurative knowledge can be extracted from the components of some idioms. By using idioms with both figurative and literal interpretations, training could mislead the student into concluding that there is a one-to-one, direct and literal link between the words comprising the idiom and the gloss of the complex phrase.

The final steps in a training program would involve introducing idioms that have plausible literal and figurative interpretations and unfamiliar as well as

familiar idioms of various forms, to train flexibility in the analytic processes used to gloss the figurative meanings.

### Limitations

There are constraints on the generalization of the results obtained in this study due to limitations inherent in subject selection, and in the choice of target stimulus items. These constraints are discussed separately, below.

Participants. The criteria used to select students for participation in the experimental tasks impose several constraints on the generalization of the results. Perhaps most critically for students with typical language development, all participants were (a) from English-only households, and (b) were overwhelmingly non-Hispanic white. These two demographics are substantially different from those of the regional and national populations. Since idioms are language and culture specific, the use of white, English-only students may have generated different patterns or levels of scores than if the sample was demographically representative. This would be particularly true for students in the earliest grades, when the balance of language experience would still tilt toward family and friends, not toward a balance with the language and literature of the majority culture.

A more technical limitation relates to the distribution of the population sampled. Specifically, selection criteria trimmed students from both tails of the normal distribution of linguistic and cognitive skills, with averaged subtest scores in each skill area limited to +/- one standard deviation around the mean. This may have reduced both the variance of, and variability among the participants' scores. Naturally, use of these criteria also means that the results are not at all representative of the responses of students with atypical language development.

Stimuli. The target idioms used are the source of two limitations. First, the use of unfamiliar idioms precludes directly integrating these outcomes with any results obtained using familiar idioms. Next, the anomalous idioms used are an eclectic collection of items, representing varied levels of familiarity and analyzability, and as such the degree to which they may represent the total corpus of anomalous idioms is entirely coincidental.

**Methodology.** The high accuracy rates obtained in item-selection, with ceiling effects apparent by G8, indicates that providing only three choices for selection may have produced results that over-represent students' idiomatic comprehension. The results may to some degree reflect 'educated guessing' by the participants, which, by eliminating the least plausible alternative from among the three choices could result in a somewhat artificial representation of what the participants 'know' about idioms.

### **Future Research**

These results did 'add into the balance of the debate on analyzability', as one early objective had hoped. However, the results also paint a complex pattern of effects and interactions, with unanticipated findings arising from the interplay of factors affecting comprehension processes. This indeed calls for yet future research into the development of idiom comprehension. Two primary recommendations for future research logically follow from the results and discussion -- for replication, and for expanding upon these outcomes -- and these are discussed next.

**Replication.** Results that contain new outcomes (the interaction of analyzability and response type; the differences in development between VP idioms and both NP and ADJP idioms), and which present both complex and specific outcomes demand replication. In replication, I would recommend using a larger initial pool of idioms for familiarity and analyzability ratings. There would be several benefits: (1) to identify groups of even less familiar idioms, with analyzability ratings representing representing yet greater extremes of SA - SU; and (2) to hopefully identify unfamiliar anomalous idioms that could be used in at least some of the preplanned portions of a study.

**Expansion.** To study idioms is to "address the concept of a lexical component in a grammatical framework" (Everaert, et al., 1995, p.2). In this dissertation, it was argued that addressing this concept within a discourse framework was also necessary, at least in developmental studies. Possibilities for future research in relation to studying comprehension within discourse arise in several areas. First, target idioms presented in context are always presented

within the terminal sentence, often as the final phrase in a short story. What might happen in comprehension if an idiom was integrated into the body of a narrative, with the narrative action carrying on beyond the idiom, rather than using the idiom as the final element? And, expanding on this suggestion, the idiomatic meaning is then usually the only focus in assessing comprehension after this 'terminal event'. If the claim for discourse processes is valid, research should also investigate how children retell narratives that contain idioms after they have heard them. Use of story recasts could provide evidence regarding how children use the figurative meanings in making sense of narratives, and perhaps of how they maintain narrative sensibility when they cannot gloss the figurative meaning.

For obvious reasons of stimulus control, idioms are usually presented individually in narratives. However, it would also be interesting to explore comprehension processes when idioms are manipulated in more complex ways. For example, what if two characters used idioms to carry dialogue forward in a story: 'I think the spy must have *kicked the bucket* yesterday' 'I don't agree, I think he's still running *hither and yon* in our classified files'. What explanations might students offer to account for this negotiation of conflict within the narrative? Finally, there is general agreement that figurative comprehension is a metalinguistically challenging activity. It might be interesting to develop short narratives in which one character specifically identifies his/her comprehension breakdown around an idiom, and invites the participant to help solve the breakdown as a part of the 'dialogue'. A story might run, 'I'm not sure I understand what you mean by *don't look a gift horse in the mouth*. I mean, yesterday you were telling me not to *buy a pig in a poke*. Can you tell me what you mean another way?' (participant responds).

Idioms may come, and idioms may go, but there's always room for investigating the core of the corpus!

## Conclusions

The term idiom is derived from the Greek *idios*, 'private'<sup>81</sup>, and can be

<sup>81</sup> Webster's II New Riverside University Dictionary (1984)

more fully be translated as “a manifestation of the unexpected” (Kilpatrick, 1998). It is fair to say that the results of the present research fit comfortably within this definition.

This study produced two useful findings. The first finding is that analyzability does impact children’s comprehension of idioms. However, the impact of analyzability is not unilateral and it was argued that analyzability’s effect can only be viewed productively from within the broader framework of discourse processes. Indeed, for at least children below high school age, consideration of the processes involved in idiom comprehension seems to require that we acknowledge and account for the multiple levels of analysis that participate in idiom comprehension.

The second finding is that the behavior of verb phrase idioms is not an adequate proxy for describing children’s comprehension of the larger corpus of idiom forms. Students are more successful in selecting the meanings of noun phrase and adjective phrase idioms than of verb phrase idioms. Yet, they are more accurate in explaining the meanings of verb phrase idioms than of the other two types. Future studies should be more broadly representative of the variety of idioms forms included in the published idiom lexicon, in part to determine whether, when, and to what degree the syntactic and semantic information contained in these complex units participates in comprehension processes.

The study of children’s comprehension of idioms is a complex and multi-faceted endeavor. One is tempted to stand on the river’s bank and take measurements of the state of the river by withdrawing samples and examining the retrieved water ever so carefully. This behavior is both useful and productive. To understand the river, however, one must examine it’s watershed, the hydraulics, and the characteristics of the terrain through which it flows. Both rivers and comprehension processes channel and braid, moving apart and even digging new channels in response to certain conditions, yet ultimately recombining, contents intact to flow into a larger body. As it is with rivers, so it seems to be as children journey on their way to an understanding of idioms.

## REFERENCES

- Abkarian, G.G., Jones, A., & West, G. (1992). Young children's idiom comprehension: trying to get the picture. Journal of Speech and Hearing Research, 35, 580 - 587.
- Abrahamsen, E., & Burke-Williams, D. (1996, November). Metaphoric transparency and syntactic frozenness in children with learning disabilities. Poster session presented at the meeting of the American Speech-Hearing-Language Association.
- Ackerman, B.P. (1982). On comprehending idioms: Do children get the picture? Journal of Experimental Child Psychology, 33, 439 -454.
- Becker, J. (1975). The phrasal lexicon. In B. Webber & R. Schank (Eds.), Theoretical issues in natural language processing. Cambridge, MA: Association for Computational Linguistics.
- Billingsley, R., White, O.R., & Munson, R. (1980). Procedural reliability: A rationale and an example. Behavioral Assessment, 2, 229 - 241.
- Boatner, M.T., Gates, J.E., & Makkai, A. (1966). A dictionary of American idioms. West Hartford, CT: American School for the Deaf.
- Boatner, M.T., Gates, J.E., & Makkai, A. (1975). A dictionary of American idioms. New York: Baron's Educational Series.
- Bobrow, S.A., & Bell, S.M. (1973). On catching on to idiomatic expressions. Memory and Cognition, 1, 343 - 346.
- Brinton, B., Fujiki, M., & Mackey, T.A. (1985). Elementary school age children's comprehension of specific idiomatic expressions. Journal of Communication Disorders, 18, 245 - 257.
- Burbules, N.C., Schraw, G., & Trathen, W. (1989). Metaphor, idiom, and figuration. Metaphor and Symbolic Activity, 4(2), 93 - 110.
- Cacciari, C., & Levorato, M.C. (1989). How children understand idioms in discourse. Journal of Child Language, 16, 387 - 405.
- Cacciari, C., & Tabossi, P. (1988). The comprehension of idioms. Journal of Language and Memory, 27, 668 - 683.

- Chafe, W.L. (1968). Idiomaticity as an anomaly in the Chomskyan paradigm. Foundations of Language, 4, 109 - 127.
- Chomsky, N. (1980). Rules and representations. New York: Columbia University Press.
- Clausner, T.C., & Croft, W. (1997). Productivity and schematicity in metaphors. Cognitive Science, 21 (3), 247 - 282.
- Cronk, B.C., & Schweigert, W.A. (1992). The comprehension of idioms: The effects of familiarity, literalness, and usage. Applied Psycholinguistics, 13, 131 - 146.
- Cutting, J.C., & Bock, K. (1997). That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. Memory & Cognition, 25 (1), 57 - 71.
- Dollaghan, C.A. (1987). Fast mapping in normal and language-impaired children. Journal of Speech and Hearing Disorders, 52(3), 218 - 222.
- Estill, R.B., & Kemper, S. (1982). Interpreting idioms. Journal of Psycholinguistic Research, 11, 559 - 581.
- Everaert, M., & van der Linden, E-J. (1989). Proceedings of the first Tilburg workshop on idioms. Tilburg: ITK.
- Everaert, M., van der Linden, E-J., Schenk, A., & Schreuder, R. (1992). In M. Everaert, E.-J. van der Linden, & A. Schenk, & R. Schreuder, (Eds.). Proceedings of IDIOMS. Tilburg: ITK.
- Everaert, M., van der Linden, E-J, Schenk, A., & Schreuder, R. (1995). Introduction (pp. 1 - 13). In M. Everaert, E. van der Linden, A. Schenk & R. Schreuder (Eds.), Idioms: Structural and psychological perspectives. Hillsdale: Erlbaum.
- Ezelle, A. (1998). A PGO action research project in language research. Paper presented at the Seattle School District C.D.S. meeting, May, 1998.
- Ezell, H.K., & Goldstein, H. (1991). Comparison of idiom comprehension of normal children and children with mental retardation. Journal of Speech and Hearing Research, 34, 812 - 819.
- Ferguson, G. A. (1981). Statistical analysis in psychology and education (5th

ed.). New York: McGraw-Hill.

- Fillmore, C.J., Kay, P., & O'Connor, M.C. (1988). Regularity and idiomaticity in grammatical constructions: The case of let alone. Language, 64 (3), 501 - 538.
- Forrester, M.A. (1994). Tropic implicature and context in the comprehension of idiomatic phrases. Journal of Psycholinguistic Research, 24 (1), 1 - 22.
- Fraser, B. (1970). Idioms within a transformational grammar. Foundations of Language, 6, 22 - 42.
- Fraser, B. (1974). The verb-particle combination in English. Tokyo: Taishukan.
- Frege, G. (1879/1972). Conceptual notation: A formula language of pure thought modeled upon the formula language of arithmetic. In T.W. Bynum (Ed. and Trans.) Conceptual Notation and Related Articles. Oxford: Oxford University Press. (First published in German, 1879; Halle: L. Nerbert).
- Gasser, M., & Dyer, M. (1986). Speak of the devil: Representing deictic and speech act knowledge in an integrated memory. Proceedings of the sixth annual meeting of the Cognitive Science Society. Hillsdale, NJ: Erlbaum.
- Gibbs, R.W. (1980). Spilling the beans on understanding and memory for idioms in conversation. Memory and Cognition, 8, 149 - 156.
- Gibbs, R.W. (1986). Skating on thin ice: Literal meaning and understanding idioms in conversation. Discourse Processes, 9, 17 - 30.
- Gibbs, R.W. (1987). Linguistic factors in children's understanding of idioms. Journal of Child Language, 14, 569 - 586.
- Gibbs, R.W. (1991). Semantic analyzability in children's understanding of idioms. Journal of Speech and Hearing Research, 34, 613 - 620.
- Gibbs, R.W. (1994). The poetics of mind: Figurative thought, language and understanding. New York: Cambridge University Press.
- Gibbs, R.W., Bogdanovich, J.M., Sykes, J.R., & Barr, D.J. (1997). Metaphor in idiom comprehension. Journal of Memory and Language, 37, 141 - 154.
- Gibbs, R.W., & O'Brien, J.E. (1990). Idioms and mental imagery: The metaphorical motivation for idiomatic meaning. Cognition, 36, 35 - 68.

- Gibbs, R.W., & Gonzales, G.P. (1985). Syntactic frozenness in processing and remembering idioms. Cognition, 20, 243 - 259.
- Gibbs, R.W., & Nayak, N.P. (1989). Psycholinguistic studies on the syntactic behavior of idioms. Cognitive Psychology, 21, 100 - 138.
- Gibbs, R.W., & Nayak, N.P. (1991). Why idioms mean what they do. Journal of Experimental Psychology: General, 120 (1), 93 - 95.
- Gibbs, R., Nayak, N.P., Bolton, J., & Keppel, M. (1989). Speakers' assumptions about the lexical flexibility of idioms. Memory & Cognition, 17, 58 - 68.
- Gibbs, R.W., Nayak, N.P., & Cutting, C. (1989). How to *kick the bucket* and not decompose: Analyzability of idiom processing. Journal of Memory and Language, 28, 576 - 593.
- Glass, A.L. (1983). The comprehension of idioms. Journal of Psycholinguistic Research, 12 (4), 429 - 442.
- Glucksberg, S. (1989). Metaphors in conversation: How are they understood? Why are they used? Metaphor and Symbolic Activity, 4 (3), 125 - 143.
- Harris, R.J. (1975). A primer of multivariate statistics. New York: Academic Press.
- Heringer, J. (1976). Idioms and lexicalization in English. In J.P. Kimball (Ed.), Syntax and semantics: The grammar of causative constructions. New York: Academic Press.
- Hoffman, R.R., & Honeck, R.P. (1980). A peacock looks at its legs: Cognitive science and figurative language. In R.P. Honeck & R.R. Hoffman (Eds.), Cognition and figurative language. Hillsdale: Erlbaum.
- Hoffman, R.R., & Kemper, S. (1987). What could reaction-time studies be telling us about metaphor comprehension? Metaphor and Symbolic Activity, 2 (3), 149 - 186.
- Honeck, R.P. (1980). Historical notes on figurative language. In R.P. Honeck & R.R. Hoffman (Eds.), Cognition and figurative language. Hillsdale: Erlbaum.
- Hollingsed, J.C. (1958). A study of figures of speech in intermediate grade reading. Unpublished doctoral dissertation, Colorado State University, Fort Collins.

- Howell, D.C. (1989). Fundamental statistics for the behavioral sciences (2nd ed.) Boston: PWS-Kent Publishing Co.
- Jackendoff, R. (1995). The boundaries of the lexicon. In M. Everaert, E-J van der Linden, A. Schenk & R. Schreuder (Eds.), Idioms: Structural and psychological perspectives (pp. 133 - 165). Hillsdale: Erlbaum.
- Jackendoff, R. (1975). Morphological and semantic regularities in the lexicon. Language, 51, 639 - 671.
- Johnson, G.A. (1985). The comprehension of idioms in normally developing and language-deficient children. Unpub. M.A. thesis, Missoula, MT.
- Katz, A.N., Paivio, A., Marschark, M. & Clark, J.M. (1988). Norms for 204 literary and 260 nonliterary metaphors on 10 psychological dimensions. Metaphor and Symbolic Activity, 3(4), 191 - 214.
- Katz, J.J., & Postal, P.M. (1963). Semantic interpretation of idioms and sentences containing them. MIT Research Laboratory of Electronics, Quarterly Progress Report, (No. 70), 275 - 282.
- Kemper, S. (1986). Inferential processing and the comprehension of idioms. Metaphor and Symbolic Activity, 1, 43 - 55.
- Kerbel, D., & Grunwell, P. (1997). Idioms in the classroom: An investigation of language unit and mainstream teacher's use of idioms. Child Language, Teaching and Therapy, 13(2), 113 - 123.
- Keysar, B., & Bly, B. (1995). Intuitions of the transparency of idioms: Can one keep a secret by spilling the beans? Journal of Memory and Language, 34, 89 - 109.
- Kilpatrick, J.J. (1998, April). Idiomatic usage adds vibrancy, texture, color to American speech. Seattle Post-Intelligencer.
- Kittay, E., & Lehrer, A. (1981). Semantic fields and the structure of metaphor. Studies in Language, 5, 31 - 63.
- Lakoff, G. (1986). The meanings of literal. Metaphor and Symbolic Activity, 1 (4), 291 - 296.
- Lakoff, G. (1987). Women, fire, and dangerous things: What categories reveal about the mind. University of Chicago Press.

- Lakoff, G., & Johnson, M. (1980). Metaphors we live by. Chicago: University of Chicago Press.
- Lazar, R.T., Warr-Leeper, G.A., Nicholson, C.B., & Johnson, S. (1989). Elementary school teachers' use of multiple meaning expressions. Language, Speech, and Hearing Services in Schools, 20, 420 - 430.
- Levorato, M.C., & Cacciari, C. (1995). The effects of different tasks on the comprehension and production of idioms in children. Journal of Experimental Child Psychology, 60, 261 - 283.
- Levorato, & Cacciari, C. (1992). Children's comprehension and production of idioms: the role of context and familiarity. Journal of Child Language, 19, 415 - 433.
- Lodge, D.N., & Leach, E.A. (1975). Children's acquisition of idioms in the English language. Journal of Speech and Hearing Research, 18, 521 - 529.
- Makkai, A. (1984). A dictionary of American idioms. New York: Barron's.
- McGlone, M.S., Glucksberg, S. & Cacciari, C. (1994). Semantic productivity and idiom comprehension. Discourse Processes, 17, 167 - 190.
- Miller, G.A., & Johnson-Laird, P.N. (1976). Language and perception. Cambridge, MA: Harvard University Press.
- Mueller, R.A., & Gibbs, R.W. (1987). Processing idioms with multiple meanings. Journal of Psycholinguistic Research, 16 (1), 63 - 81.
- Nayak, N.P., & Gibbs, R.W. (1990). Conceptual knowledge in the interpretation of idioms. Journal of Experimental Psychology: General, 119, 315 - 330.
- Newmeyer, F. (1972). The insertion of idioms. In Papers from the eighth regional meeting of the Chicago Linguistics Society (pp. 294 - 302). Chicago: Chicago Linguistics Society.
- Newmeyer, F. (1974). The regularity of idiom behavior. Lingua, 34, 327 - 342.
- Nippold, M.A. (1991). Evaluating and enhancing idiom comprehension in language - disordered students. Language, Speech and Hearing Services in Schools, 22, 100 - 106.

- Nippold, M.A., & Martin, S.T. (1989). Idiom interpretation in isolation versus context: A developmental study with adolescents. Journal of Speech and Hearing Research, 32, 59 - 66.
- Nippold, M.A., & Rudzinski, M. (1993). Familiarity and transparency in idiom explanation: A developmental study of children and adolescents. Journal of Speech and Hearing Research, 36, 728 - 737.
- Nippold, M.A., & Taylor, C.L. (1995). Idiom understanding in youth: Further examination of familiarity and transparency. Journal of Speech and Hearing Research, 38, 426 - 433.
- Nippold, M.A., Taylor, C.L., & Baker, J.M. (1996). Idiom understanding in Australian youth: A cross-cultural comparison. Journal of Speech and Hearing Research, 39, 442 - 447.
- Nunberg, G. (1978). The pragmatics of reference. Indiana University Linguistics Club.
- O'Brien, R.G., & Kaiser, M.K. (1985). MANOVA method for analyzing repeated measures designs: An extensive primer. Psychological Bulletin, 97 (2), 316 - 333.
- Ortony, A. (1980). Some psycholinguistic aspects of metaphor. In R.P. Honeck & R.R. Hoffman, (Eds.), Cognition and figurative language. Hillsdale, N.J.: Erlbaum.
- Ortony, A., Schallert, D.L., Reynolds, R.E., & Antos, S.J. (1978). Interpreting metaphors and idioms: Some effects of context on comprehension. Journal of Verbal Learning and Verbal Behavior, 17, 465 - 477.
- Oxford English Dictionary (2nd ed.)(1933) Oxford: Clarendon Press.
- Paivio, A. (1979). Psychological processes in the comprehension of metaphor. In A. Ortony, (Ed.), Metaphor and thought. New York: Cambridge University Press.
- Perrine, L. (1971). Four forms of metaphor. College English, 33, 125 - 138.
- Popiel, S.J., & McRae, K. (1988). The figurative and literal senses of idioms, or all idioms are not used equally. Journal of Psycholinguistic Research, 17 (6), 475 - 487.
- Prinz, P. M. (1983). The development of idiomatic meaning in children.

- Reagan, (1987). The syntax of English idioms: Can the dog be put on? Journal of Psycholinguistic Research, 16 (5), 417 - 441.
- Richards, I.A. (1936). Metaphor. In I.A. Richards, (Ed.), The Philosophy of Metaphor. London: Oxford University Press.
- Schenk, (1995). The syntactic behavior of idioms. In M. Everaert, E.-J. van der Linden, A. Schenk & R. Schreuder (Eds.), Idioms: Structural and psychological perspectives. Hillsdale: Erlbaum.
- Schraw, G., Trathen, W., Reynolds, R.E., & Lapan, R.T. (1988). Preferences for idioms: Restrictions due to lexicalization and familiarity. Journal of Psycholinguistic Research, 17 (5), 413 - 424.
- Schweigert, W.A. (1991). The muddy waters of idiom comprehension. Journal of Psycholinguistic Research, 20 (4), 305 - 314.
- Schweigert, W.A. (1986). The comprehension of familiar and less familiar idioms. Journal of Psycholinguistic Research, 15, 33 - 45.
- Schweigert, W.A., & Moates, D.R. (1988). Familiar idiom comprehension. Journal of Psycholinguistic Research, 17 (4), 281 - 296.
- Seidl, J., & McMordie, W. (1978). English idioms and how to use them. Hong Kong: Oxford University Press.
- Semel, E., Wiig, E. & Secord, W. (1987). Clinical Evaluation of Language Fundamentals-Revised. San Antonio: The Psychological Corporation, Harcourt Brace & Company.
- Stark, R. E., & Tallal, P. (1981). Selection of children with specific language deficits. Journal of Speech and Hearing Disorders, 46, 114 - 122.
- Strand, K., & Fraser, B. (1979, October). The comprehension of verbal idioms by young children. Paper presented at the Boston University Language Development Conference, Boston, MA.
- Swinney, D.A., & Cutler, A. (1979). The access and processing of idiomatic expressions. Journal of Verbal Learning and Verbal Behavior, 18, 523 - 534.

- Thorum, A.R. (1986). The Fullerton Language Test for Adolescents (2nd ed.). Palo Alto CA: Consulting Psychologists Press.
- Van de Voort, M.E.C., & Vonk, W. (1995). You don't die immediately when you kick an empty bucket: A processing view on semantic and syntactic characteristics of idioms. In M. Everaert, E.-J. van der Linden, A. Schenk & R. Schreuder (Eds.), Idioms: Structural and psychological perspectives. Hillsdale: Erlbaum.
- Van Lancker, D., & Canter, G. (1981). Idiomatic versus literal interpretation of ditropically ambiguous sentences. Journal of Speech and Hearing Research, 24, 523 - 34.
- Wasow, T., Sag, I., & Nunberg, G. (1983). Idioms: An interim report. In S. Hattori & K. Inoue, (Eds.), Proceedings of the XIIIth International Congress of Linguistics. Tokyo: CIPL.
- Webster's II New Riverside University Dictionary (1984). Boston: Houghton Mifflin Co.
- Wechsler, D. (1974). Wechsler Intelligence Scale For Children-Revised. New York: The Psychological Corporation.
- Weinreich, U. (1969). Problems in the analysis of idioms. In J. Puhvel (Ed.), Substance and structure of language. Berkeley: University of California Press.
- Wiig, E., & Secord, W. (1992). Test of Word Knowledge. The Psychological Corporation: Harcourt Brace Jovanovich, Inc.

## Appendix A: Instructions for Idiom Analyzability Ratings

### Idiom Ratings: Meanings

In this experiment your task is to rate how directly the figurative meaning of each idiom relates to the paraphrase associated with it. That is, do the words that comprise the idiom help you 'figure out' the idiomatic meaning?

Examples of idioms in which the words do bear a relatively direct relationship to the figurative meaning of the idiom include: (a) *go to waste* (to be unused); (b) *lose your grip* (to lose control); and (c) *close your eyes to* (ignore something). Idioms like these are called 'transparent'; you can in a sense 'see through' the surface form to the underlying meaning.

For other idioms, there is no obvious relationship between the words that comprise the idiom and the idiom's figurative meaning. The words that comprise the idiom are not useful in helping you 'figure out' the idiomatic meaning. Examples of idioms in which the words bear no direct relationship to the figurative meaning of the idiom include: *kick the bucket*, "to die"; *fall off the wagon*, "to start drinking again"; and *face the music*, "to take responsibility". Idioms like these are called 'opaque' (Gibbs & Nayak, 1989); like a frosted window, you cannot 'see through' the surface form to the underlying meaning.

On the following pages a variety of idioms are given, followed by a paraphrase of the idiom's meaning, in parentheses. Please read each idiom and its paraphrase, then rate each idiom as TRANSPARENT or OPAQUE, along this 5 - point continuum:

- 1) No relationship between the idiom's figurative meaning and the words that make up the idiom (as in *kick the bucket*, meaning "to die");
- 3) There is a partial or indirect relationship between the idiom's figurative meaning and the words that make up the idiom (as in *bury the hatchet*, meaning "to resolve a dispute");
- 5) There is a direct relationship between the idiom's figurative meaning and the words that make up the idiom (as in *skating on thin ice*, meaning "to risk danger, disapproval, or anger").

131

Take as much time as you like to think about each item. Please do not go back and change any item. Please try to do this list in one pass.

Put an **X** through the number you select.

You will be paid \$7.00 for your time/effort; please remember to sign this sheet so I know who I owe money to!

## Appendix B: Instructions for idiom familiarity ratings

### Idiom Ratings

**Instructions to each Judge:** Please read the following idiomatic phrases, and rate how familiar each idiom is to you. Please rate each idiom immediately after reading it; do not spend time trying to 'think through' each item. Your immediate reaction is what I am interested in. Please do not go back and change or compare any of your responses. This really isn't a test in any sense. There are no 'right' or 'wrong' answers!

#### **Rating Scale:**

1. Uncommon; you've either never heard/read this phrase, or perhaps have encountered it once or twice.
2. Somewhat uncommon; you're pretty sure you've heard/read this phrase on several occasions, but its not used in literature or conversation on a regular basis.
3. Somewhat common; maybe be used by particular people groups, or encountered in particular kinds of reading material.
4. Common; in everyday use, in literature and/or conversation.

Put an **X** through the number you select.

Please do this list in one pass.

You will be paid \$7.00 for your time/effort; please remember to sign this sheet so I know who I owe money to!

## Appendix C: Stimuli

Target Idioms, associated short stories and forced choice paraphrases.

### Verb Phrase idioms (VP) - Semantically analyzable (SA):

1. go to bed with the chickens

Ginger was certainly a hard worker. She had lots to do all day every day, and she always had to be alert. We hardly ever saw her. She got up early each day and *went to bed with the chickens*.

- p<sup>1</sup> – to go to bed early at night
- l – to sleep with domestic fowl
- io – to eat out of his/her hand

2. meet your Waterloo

The Jets won 9 playoffs games in a row. Everybody thought they would repeat as champs. Their coach warned them not to be too confident. He said, if you're not careful, the Eagles will make sure we *meet our Waterloo*.

- p – to lose an important contest
- l – to go to the town, "Waterloo"
- io – to go fly a kite today

3. lord it over

Everybody said Jason was a real jerk. He was rich, good-looking, and really arrogant. I never liked him at all. One day, I went up to him and told him that he better not try to *lord it over* me.

- p – to act as the superior and master of
- l – to be given a title of royalty
- io – to take leave of someone

4. ask for the moon

Simon always thought well of himself. He worked hard and he knew what he was doing. When he told me about the raise he asked the boss for, I said, boy Simon, you really *asked for the moon*.

- p – to want something you cannot reach
- l – to request the earth's natural satellite
- io – to save your breath for later

---

<sup>1</sup> Legend: p = accurate paraphrase of the figurative meaning; l = literal paraphrase; io = accurate paraphrase of a different idiom, not used as a stimulus item

Verb Phrase (VP) - Semantically Unanalyzable (SU)

## 1. carry coals to New Castle

Steve bought a case of Pepsi for the party. He'd forgotten that Annika already had two cases. When he walked in with the pop and saw all the other cans, he said boy, this is like *carrying coals to New Castle*.

paraphrase (p) – to do something unnecessary

literal (l) – to bring fuel to a town

other idiom (io) – to go down in history

## 2. fly in the face of

The street was busy, but Phil was in a hurry to get home, so he crossed against the light. When his mother found out she was furious. She said to him this *flies in the face of safety*.

p – to ignore or go against something

l – to strike someone's face while airborne

io – to fly by the seat of your pants

## 3. ring the changes

Erin wanted a motorbike badly, but her parents wouldn't buy her one. She left out ads, talked with her brothers about it, and told her parents she was tired of riding the bus. Finally, her parents got frustrated and told her to stop *ringing the changes*.

p – to say the same idea in different ways

l – to celebrate some alterations

io – to land on your feet

## 4. give to understand

Maria was really excited; she thought that class was canceled for tomorrow. I told her no, that class would meet as usual. Maria said, 'how funny, that's what Freda had *given me to understand*.

p – to make somebody think something is true without telling them

l – to share knowledge with someone

io – to raise the roof over your head

Noun Phrase (NP) - Semantically Analyzable (SA)

## 1. The horns of a dilemma

The boss told Angelina that she had to fire either her secretary or her receptionist. Angelina knew the company was losing money, but she didn't want to fire anybody. She found herself on the *horns of a dilemma*.

p – two choices possible, with neither choice wanted  
 l – atop the antlers of a magical creature  
 io – the ins and outs of something

## 2. A bee in her bonnet

June thought that she knew a shortcut home. Steve didn't think she was right, and they discussed her plan for a long time. Steve told her, 'June, you've got a real *bee in your bonnet*.

p – a fixed idea that seems fanciful, odd, or crazy  
 l – an insect trapped in a hat  
 io – to get a fix on that idea

## 3. The fat of the land

Crossing the great plains in covered wagons was hard and dangerous. Many settlers were unprepared for the hardships, and they suffered terribly. Mistakenly, many thought that they could live off of *the fat of the land*.

p – the best and richest of food, clothes, everything  
 l – a region's excess production  
 io – where the shoe pinches the foot

## 4. A law unto one's self

The young cowboy was proud and quick with his six-shooter. The townspeople made sure that they stayed out of his way. They knew all too well that he was *a law unto himself*.

p – a person who does only as they wish  
 l – someone who maintains the peace  
 io – to know a thing or two

Noun Phrase (NP) - Semantically Unanalyzable (SU)

## 1. A straw in the wind

Last Friday I noticed that the faucet was dripping. I tightened the washer, and the drip stopped. On Monday, the faucet exploded. Only then did I realize that the drip was *a straw in the wind*.

p – a small sign of what may happen  
 l – dry grass blowing in the breeze  
 io – a feather in your cap

2. A dressing down

Carol's mom asked her to turn off the hose after she'd watered the garden. She forgot, and the water police fined her mom \$100. When Carol came home later her mother gave her *a real dressing down*.

p – a scolding  
 l – taking clothes off over the feet  
 io – to have irons in the fire

3. A pig in a poke

Ted worked all summer to buy his first car. When he finally saved enough, he got so excited he bought the first car he saw without even driving it. His dad told him, well son, you've bought yourself *a pig in a poke*.

p – something bought without careful examination  
 l – a farm animal in a sack  
 io – to have bats in the belfry

4. The life of Riley

Jake's mother looked after the children, cooked, cleaned, and worked full time. Jake's dad worked part time, but at home all he did was watch TV and drink beer. Jake's grandma said his dad was living the *life of Riley*.

p – a soft, easy life  
 l – the life of a man named 'Riley'  
 io – to know what is what

Adjective Phrase (ADJP) - Semantically Analyzable (SA)

1. Behind the eight-ball

Alex promised to complete the assignment by Monday. He put it off until Sunday night, and when he switched his computer on he saw his hard disk had crashed. He said, boy, am I ever *behind the eight-ball*.

p – in a difficult position  
 l – to the rear of the black ball in pool  
 io – after your own heart

## 2. Out of the swim

Amy's divorce after 10 years of marriage had been long and very messy. Her friend asked why she hadn't started dating again. She said, I don't know, I guess that I'm just *out of the swim*.

- p – not as socially active as others are
- l – no longer propelling yourself through the water
- io – too big for your britches

## 3. Bred in the bone

Leroy found a gym bag filled with money in a deserted alley. His friends thought he was crazy when he turned it over to the police. When asked why he did that, he said that it was just *bred in the bone*.

- p – belonging to your nature or character
- l – born into part of the skeleton
- io – really for the best

## 4. Asleep at the switch

The game was tied, but the bases were loaded. Unexpectedly, Kevin watched as strike three came in belt high right across the plate. Later the manager said that Kevin was *asleep at the switch*.

- p – failing to act promptly, as expected;
- l – dormant by a circuit selector
- io – definitely out of shape

Adjective Phrase (ADJP) - Semantically Unanalyzable (SU)

## 1. Beyond the pale

Alice and Jean had been friends for many years. One day, Jean invited Alice's boyfriend to a concert without telling Alice. Alice was furious when she found out. She said that Jean was now *beyond the pale*.

- p – in disgrace, with no chance of being accepted
- l – outside of the English domains
- io – completely under the weather

## 2. All wool and a yard wide

Allen was always good to his family and friends. He gave to charity and did volunteer work each weekend at the shelter. His friends said that he was *all wool and a yard wide*.

- p – of fine character, very generous and kind-hearted
- l – composed of sheep's fur, three feet across
- io – just totally made of money

## 3. Three sheets to the wind

Late one night Leon came home from the party and started pounding on our door. When I let him in he tripped over the rug, fell down, and started giggling. I said "Leon, you're *three sheets to the wind*."

p – unsteady from too much liquor  
l – several bed coverings waving in the air  
io – as large as life

## 4. Fed to the teeth

It was a rainy Sunday afternoon and the children were bored. They were in to everything and they never shut up for even a second. Finally I told them that their behavior had me *fed to the teeth*.

p – at the end of your patience  
l – no room for any more food  
io – entirely for the birds

Anomalous Idioms - Unrated for Analyzability

## 1. Come a cropper

The district playoffs were in two weeks. John was so confident of winning that he stopped training. His coach was worried. He told John that if he didn't train he would *come a cropper*.

p – to fail at something  
l – to fall off of a horse  
io – to be safe and sound

## 2. Tit for tat

This morning, Dave swiped Bill's good pen from his desk. That afternoon, Bill stole Dave's baseball cap from his locker. Bill thought to himself, "there, that's *tit for tat*".

p - equal treatment in return  
l – to rap with light blows  
io – to know what is what

## 3. By and by

"I'll die if I can't go to Disneyland, I told my parents. They said that we didn't have the time to take a trip this summer. When they realized how sad I was, they said, we'll get to Disneyland *by and by*."

p – at some time in the future  
l – very close to  
io – off and on

## 4. By and large

I was crushed to hear the White Sox had deliberately lost the World Series for some gamblers. I told dad I would never believe in baseball again. He told me to calm down, and said that the game was still honest, *by and large*.

p – as most often happens

l – next to and of great size

io – well and good

## 5. Hither and yon

The children were excited to visit Kelsey Creek Farms in the springtime. John went to go see the rabbits. As he opened the gate to the hutch, the rabbits all ran *hither and yon*.

p – in one direction, and then another

l – toward this place, and that one over there

io – thus and so

## 6. Which is which

Caves contain many unusual rock formations. Our teacher told us that stalactites grow down from the ceiling and stalagmites up from the floor. However, I can never recall *which is which*.

p – figure out who/what is one thing and what the other

l – chose the choice

io – enough is enough

## 7. Less than no time

We were totally depressed. The last ferry was leaving in 30 minutes, and our car had just broken down. We were overjoyed when the mechanic told us he would have it fixed in *less than no time*.

p – very quickly

l – before you begin

io – once and for all

## 8. Catch as catch can

The painters were finished for the day, but the kitchen was a mess. Everyone was hungry, and we bugged mom to make dinner. She said, I'm sorry, but tonight its *catch as catch can*.

p – in any way possible

l – positively grab a hold of

io – for as much as

Training Items

## 1. Given the bounce (VP - SU)

I went to visit Gil, and found that he was very sad. I asked him why he was so sad. He said that, to his surprise, his girlfriend had just "given him the bounce".

p- to lose one's friend or lover; to lose a job

l- to catch an elastic rebound

io- to rain cats and dogs

## 2. A case in point (NP - SA)

Ray tried to convince Al that smoking cigarettes was unhealthy. Ray said that Mrs. Downs had just died from lung cancer after smoking for 30 years, and she was only 50 years old. Al said that was different, but Ray said no, it was a "case in point".

p - an example that proves something or makes it clearer

l - where good knives are displayed

io - all for the best

## Appendix D: Instructions to the Participants

### Session One

#### General Introduction:

"First, I want to thank you for volunteering to help with this project. I am here because I'm interested in learning more about how school children's understanding and use of idioms develops. Idioms are funny parts of language; you say one thing, but mean another. For example, *kick the bucket* is an idiom. While it might mean "hit a bucket with your foot", we often understand it to mean "to die". For example, "I heard that old Mr. Jones finally *kicked the bucket* yesterday" means "Mr. Jones died yesterday".

I am interested in idioms for several reasons. First, they are used a lot in language, but we don't really know how children come to understand and use them. Because they come to be used a lot in everyday conversation and because you will run across many unfamiliar idioms over the years as you read, knowing how they are usually learned is important in education. Happily, most school children learn idioms without any special effort. The other reason I've asked for your help is that understanding and using idioms seems to be much harder for children who are having trouble learning to speak or learning to read, and for people who are learning English as a second language. In order to help these people as much as we can, we need to first know how children in regular education come to know idioms. That's why I've asked for your help.

The work that we are going to do will be recorded onto this tape recorder. This is so my work can be checked by someone else, to see that I am doing everything just right. Also, the taping will let me go back and listen carefully to your answers. I will see you two times: today, and again in a week or two for some follow-up work. Do you have any questions?"

#### Selection Measures:

"Today, we will work for 30 - 45 minutes. Today's work is a survey of your listening and speaking skills in general. If, once we get started, you decide that you really don't want to do this work, just let me know. You can leave immediately and return to class, and that will be fine with everybody. There are

specific instructions for each part that I will read to you, and there are always practice items. I won't be able to tell you how you're doing, because I won't know now. Feel free to ask questions, but remember that sometimes the rules tell me that I won't always be able to answer them. Are you ready?"

#### Session Two: Experimental measures

"Today we're going to listen to some short stories. After I read each story, I will ask some questions about that story. Some of the questions will be easy to answer, while others might be a little more difficult. Try your best in answering each question, but don't worry about whether a particular answer is 'right', or 'wrong'. There are no 'right' or 'wrong' answers; I'm just interested in what you know about each story. Remember, if at any time you don't want to continue, just tell me and you can go back to your classroom. Are you ready to begin?"

"This is what we will do with each story. I will read the story, and when I'm finished, I will repeat the last sentence in the story and ask you to tell me what it means to you. After you've done that, I will read you three different sentences, and your job will be to pick the one that means the same thing as the last sentence in the story. Let's try a story for practice. Are you ready? (First story is read to child). Good listening. Now, let's try another."

If the subject shows by his/her response that the task has not been understood, the investigator responds with: "Nice try. Let's listen to that story again, and see that else it might mean. (Subject responds) Good work. Now we are going to listen to some more stories. I will not be able to answer any questions or explain any of the stories to you. Remember to listen carefully. Are you ready?" If an answer is unclear or ambiguous, the investigator can use one prompt ("Tell me a little more"; or, I didn't quite get that, could you explain some more"). One repetition of the story and/or of the three forced-choice sentences is permitted, at the subject's request.

## VITA

Glenn A. Johnson

### EDUCATION

1998, Ph.D., Speech and Hearing Sciences, University of Washington

1985, M.A., Communication Sciences and Disorders, University of Montana

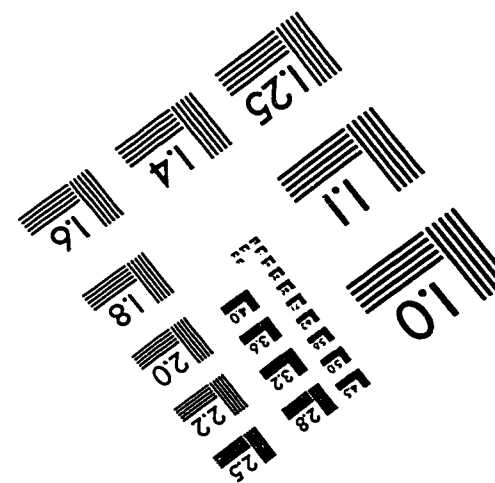
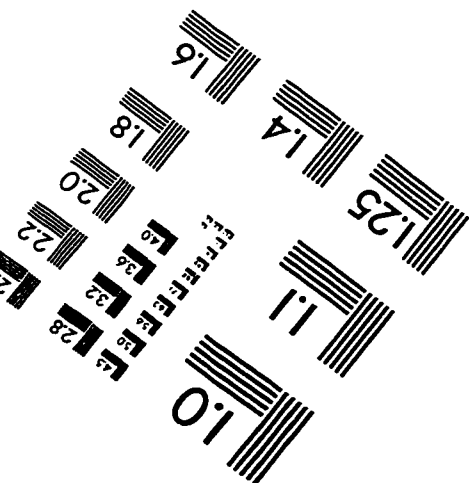
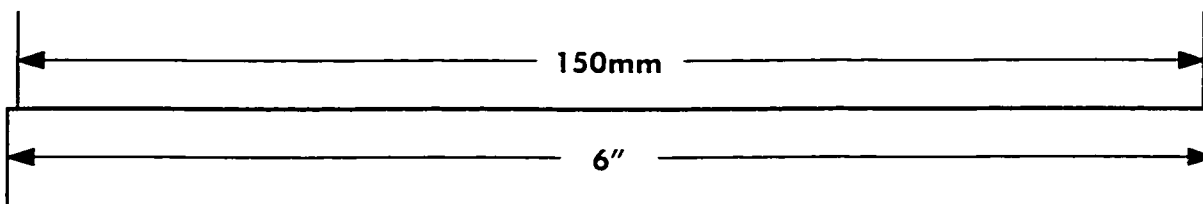
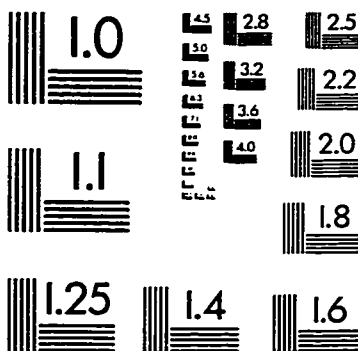
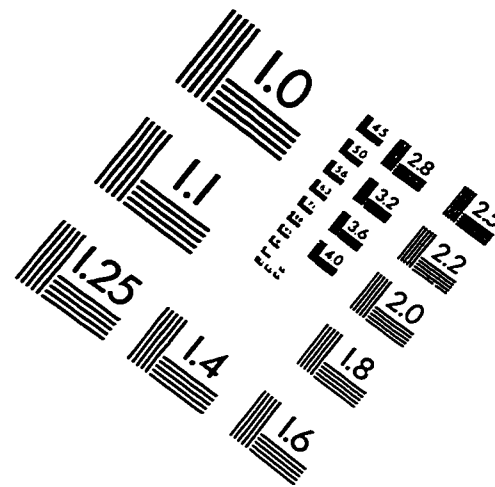
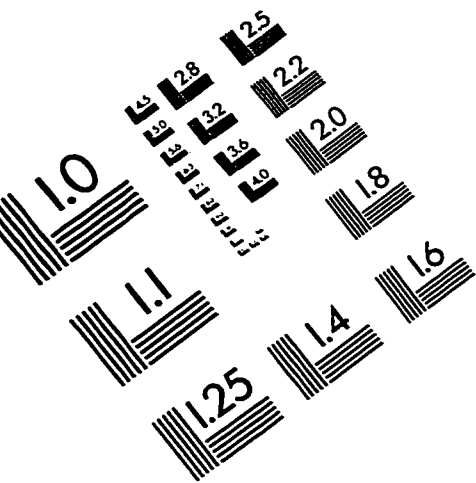
1976, B.A., Speech, Temple University

### PUBLICATIONS

Olswang, L.B., Bain, B.A., & Johnson, G.A. (1992). Using dynamic assessment with children with language disorders. In S.F. Warren and J. Reichle, (Eds.), Causes and effects in communication and language intervention. Baltimore, MD: P. Brookes, Pub..

Bain, B.A., Olswang, L.B., & Johnson, G.A. (1992). Language sampling for repeated measures with language-impaired pre-schoolers: A comparison of two procedures. Topics in Language Disorders, 12 (2), 13 - 27.

# IMAGE EVALUATION TEST TARGET (QA-3)



**APPLIED IMAGE, Inc**  
 1653 East Main Street  
 Rochester, NY 14609 USA  
 Phone: 716/482-0300  
 Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved