Title
Comparatively speaking : a psycholinguistic study of optionality in grammar

Author
Boyd, Jeremy Kenyon

Publication Date
2007

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Comparatively speaking: A psycholinguistic study of optionality in grammar

A Dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in

Linguistics and Cognitive Science

by

Jeremy Kenyon Boyd

Committee in charge:

Professor Farrell Ackerman, Chair
Professor Jeff Elman
Professor Victor Ferreira
Professor Roger Levy
Professor Rachel Mayberry

2007

The Dissertation of Jeremy Kenyon Boyd is approved, and it is acceptable in quality and form for publication on microfilm:

Chair

University of California, San Diego

2007

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

VITA

Education

2007	Ph.D., Linguistics and Cognitive Science University of California, San Diego
1999	M.A., Teaching English as a Second Language University of Idaho
1997	B.A., English Literature Magna Cum Laude University of Idaho

Appointments

2007	Postdoctoral Researcher (with Adele Goldberg) Princeton University Department of Psychology
2004-2006	Predoctoral Fellow UCSD Center for Research in Language
2001-2007	Teaching Assistant UCSD Department of Linguistics
2000-2006	Teaching Assistant UCSD Human Development Program
2003	Teaching Assistant UCSD Department of Cognitive Science
2000-2002	Teaching Assistant UCSD Humanities Program
1997-1999	Teaching Assistant University of Idaho Department of English

Talks & Publications

Boyd, J. & Slevc, L.R. (in prep). Optionality in comparative production.

Boyd, J. & Ackerman, F. (in prep). A word-based critique of Embick (2007).

Boyd, J. (2007). Comparatively speaking: A psycholinguistic study of optionality in grammar. Ph.D. thesis, University of California, San Diego.

Boyd, J. (2006). Two stages of overgeneralization in English comparative production. Poster at the *Boston University Conference on Language Development*. Boston, MA. November 4.

Boyd, J. (2006). Perseveration in English comparative production. *Center for Research in Language Colloquium.* UCSD, La Jolla, CA. October 24.

Boyd, J. (2006). On the representational status of /s/-clusters. *San Diego Linguistics Papers*, 2, 39-84. Available online at http://repositories.cdlib.org/ucsdling/sdlp2/4.

Slevc, L.R. & Boyd, J. (2006). Why it's better to be *more hungry* than *hungrier*: Optionality in comparative production. Poster at the *CUNY Conference on Human Sentence Processing*. New York, NY. March 23.

Boyd, J. (2005). Syntactic determinants of comparative form. *Linguistics Department Colloquium*. UCSD, La Jolla, CA. October 10.

Ackerman, F. & Boyd, J. (2005). Hungarian morphology and Williams Syndrome: A word-based morphological and neuroconstructivist perspective. *International Symposium on Creativity, Mind, and Brain in Hungarian Scholarship: Past and Present*. Indiana University. Bloomington, IN. April 3.

Boyd, J. (2005). A processing advantage for analytic comparatives in English. *Center for Research in Language Colloquium*. UCSD, La Jolla, CA. March 1.

Boyd, J. (2004). Comparative adjective acquisition in Williams Syndrome. *Center for Human Development Colloquium*. UCSD, La Jolla, CA. January 9.

Boyd, J. (2003). Natural language suppletion: Memory limitations and typology. *Center for Research in Language Colloquium*. UCSD, La Jolla, CA. February 4.

Boyd, J. (2003). Computational limits on natural language suppletion. *Annual Meeting of the Linguistic Society of America*. Atlanta, GA. January 4.

Boyd, J. (2002). Results from a GLA model of onset acquisition: Do English /s/-clusters differ structurally from other consonant sequences? *Southwest Workshop on Optimality Theory/Texas Linguistic Society*. Austin, TX. March 1.

Boyd, J. (1997). Words as weapons: Macrocognition, microcognition, and metaphor. *The Banks Awards*, 1996-1998, 107-113.

Awards

NSF Dissertation Improvement Grant 0617710: Optionality in Comparative Inflection. July 2006-July 2008.

NIH Training Grant 5-T32-DC0041: Language, Communication and the Brain. September 2004-August 2006.

ABSTRACT OF THE DISSERTATION

Comparatively speaking: A psycholinguistic study of optionality in grammar

by

Jeremy Kenyon Boyd

Doctor of Philosophy in Linguistics and Cognitive Science

University of California, San Diego, 2007

Professor Farrell Ackerman, Chair

GRAMMATICAL OPTIONALITY refers to the ability to realize the same meaning using more than one grammatical expression, as in the dative alternation, or the choice between active and passive encodings of a transitive event. Optionality has been studied extensively in the syntactic domain, and has provided numerous insights concerning the kinds of factors that influence production, the functional benefits of speaker choice, and the nature of syntactic development. In the present work, I inaugurate the psycholinguistic study of morphological optionality by considering an English alternation that occurs relatively commonly among adjectives in the comparative, namely the ability to inflect SYNTHETICALLY (e.g., *angrier*, *prouder*), or ANALYTICALLY (e.g., *more angry*, *more proud*). This investigation is motivated based on the pervasive character of morphological optionality cross-linguistically, and the fact that the English comparative has been unjustifiably used by a number of analysts as evidence in favor of a phenomenon known as BLOCKING. Blocking analyses view inflectional class as a

categorical matter. I demonstrate that this has the effect of erroneously predicting that optionality should not occur.

After a preliminary introduction in Chapter 1 to the issues and phenomena under investigation, I proceed in Chapter 2 to a series of experiments on comparative comprehension. Using both offline (Experiments 1 and 3) and online (Experiment 2) methods, I show that, for adjectives that participate in comparative optionality, listeners roundly prefer the analytic variant, most likely because it allows for earlier phrase-structure recognition, and avoids a temporary parsing ambiguity associated with the synthetic. Chapter 3 is a production study conducted jointly with L. Robert Slevc. In this work, we use an elicited production technique (Experiment 3) to establish that speakers tend to increase their rate of use of the analytic variant under conditions of increasing syntactic and semantic complexity. We argue that there is no evidence to suggest that this is done as a means of helping lessen addressee processing loads. The behavior is, however, consistent with the notion that speakers choose between competing variants as a means of facilitating their own production processes. In Chapter 4, I consider how the analytic and synthetic patterns of comparative inflection are acquired. Experiment 4 demonstrates that, contra the results of previous work on comparative development, both patterns are subject to overgeneralization errors (e.g., **dangerouser*, **more fast*, cf. *faster*). The results of Experiment 5 suggest that the discrepancy between my findings and those of other researchers is likely due to their failure to adequately control for response perseveration.

I conclude in Chapter 5 by providing an extensive critique of various blocking analyses of the English comparative. As an alternative, I suggest a gradient, pattern-based approach that is consistent with the assumptions of WORD AND PARADIGM morphology. This analysis avoids becoming mired in difficulties that afflict more syntactocentric treatments of comparative inflection, and allows categorical inflectional outcomes and optionality to be modeled in the same unified framework.

1. Introduction

The present work explores and analyzes GRAMMATICAL OPTIONALITY as attested in English comparative adjective inflection. This phenomenon is exemplified by simple contrasts such as *angrier* ~ *more angry*, and *prouder* ~ *more proud*. Optionality refers to the possibility of realizing the same semantic content by means of several otherwise competing grammatical expressions. In this introductory chapter, I argue that the study of optionality in the morphological domain has been largely neglected, and that further neglect is both unwise and unwarranted. In the domain of syntax, the study of optionality has led to a plethora of results that have provided significant insights into the nature of linguistic competence and processing. Given the widespread nature of morphological optionality, there is good reason to hypothesize that focused study on it will similarly advance our understanding of language. The results in the thesis, I believe, support this hypothesis.

1.1 Syntactic Optionality

In the psycholinguistic literature, the discussion of optionality has largely centered around three English alternations: the optional use of *that* to introduce sentential complements and relative clauses (2000; V. S. Ferreira & Firato, 2002; Haywood, Pickering, & Branigan, 2005; Jaeger, 2006; Roland, Elman, & Ferreira, 2006; Temperley, 2003), the *dative alternation* (Bock & Griffin, 2000; Bresnan, Cueni, Nikitina, & Baayen, 2007; Chang, Dell, & Bock, 2006; V. S. Ferreira, 1996; Huttenlocher, Vasilyeva, & Shimpi, 2004; Slevc, 2007), and *transitive voice alternations* (Bock & Griffin, 2000;

Chang et al., 2006; Huttenlocher et al., 2004; Savage, Lieven, Theakston, & Tomasello, 2003, 2006). Examples of these construction types are given below.

(1) Optional *that* mention

a. Sentence Complements: The voters believed (that) the senator truly cared.

b. Relative Clauses: The last meal (that) the prisoner ate consisted of steak, mashed potatoes, and green beans.

(2) Dative alternation

a. Prepositional Dative: The court sent [$_{NP}$ a summons][$_{PP}$ to the witness].

b. Double-Object Dative: The court sent [$_{NP}$ the witness][$_{NP}$ a summons].

(3) Transitive voice alternation

a. Active Voice: [$_{NP}$ The dog] bit [$_{NP}$ the mailman].

b. Passive Voice: [$_{NP}$ The mailman] was bitten [$_{PP}$ by the dog].

The sentences in (1) demonstrate that the function word *that* can be optionally omitted in sentence-complement constructions (1a), and in relative clauses (1b). In (2), the theme (*summons*) and recipient (*witness*) arguments of the ditransitive verb *send* can be realized as either an NP theme followed by a PP recipient (2a), or an NP recipient followed by an NP theme (2b). Likewise, the agent (*dog*) and patient (*mailman*) arguments of the transitive verb *bite* can be alternatively realized as an NP agent in subject position and an NP patient in object position (3a), or by an NP patient in subject position, and a patient that is realized as an oblique PP (3b).

Rigorous focus on these three types of alternations in English has led to important discoveries concerning the kinds of factors that influence production (Bock & Griffin, 2000; Bresnan et al., 2007; Jaeger, 2006; Roland et al., 2006), the functional benefits of speaker choice (V. S. Ferreira, 1996; V. S. Ferreira & Dell, 2000; V. S. Ferreira & Firato, 2002; Slevc, 2007), and the nature of syntactic development (Huttenlocher et al., 2004; Savage et al., 2003, 2006). These studies suggest the high potential value of exploring optionality in other languages, as well as exploring it in domains of grammar other than syntax. Note that the alternations given in (1-3) are all arguably *syntactic* in nature. Dative and active-passive voice alternations, for example, deal with the different ways in which verbal arguments can be syntactically realized. Likewise, optional *that* can be conceived of as a choice between competing syntactic structures—one in which a node for a marker of syntactic subordination is present, and one in which it is not. This primary focus on syntactic alternations, however, overlooks optionality as it occurs in other generative components of grammar—in particular, within morphology. MORPHOLOGICAL OPTIONALITY has received very little attention in the processing and acquisition literature. This neglect is particularly striking given that many morphological alternations allow for the same meaning to be mapped onto multiple forms ways that far outstrip the examples given in (1-3) in terms of formal variability.

1.2 Morphological Optionality

Inflectional morphology—the domain of investigation for the present work—is a system that formally realizes grammatical categories such as definiteness, person, number, gender, animacy, case, tense, aspect, mood, voice and polarity. Some languages mark many of these categories and are said to have rich systems of morphology. Others—like English—are relatively impoverished. Even impoverished languages, however, are capable of morphological optionality. Pinker (1999), for instance, comments on the competing regular and irregular forms that are sometimes available for realizing the past tense of English verbs such as *dive*, *smite*, *stride*, *strive* and *thrive* by saying that speakers tend to recognize their irregular realizations in books, but are often "muzzy" about whether to use them in speech. They are "tempted to default to regular forms like *smited*, *strided*, *strived*, and *thrived*. Sometimes strong [irregular] and weak [regular] forms live side by side in a person's mind, forming doublets like *strove* and *strived* or *dove* and *dived*" (p. 69). Alternations such as these constitute instances of morphological optionality. In each case, a stem bearing lexical-semantic meaning is associated with a morphosyntactic feature set (consisting solely of tense in these examples). This complex of lexical and grammatical meanings can be mapped onto either of two forms: one that follows the -*ed* pattern of past tense inflection, or one that is irregular.

In languages with rich systems of morphology, the realizational possibilities are often multiplied beyond two choices. Such languages provide striking evidence of the prodigious nature of morphological optionality: they illustrate instances of variability that are sometimes significantly greater than what is evident in the syntactic alternations shown in (1-3). Chintang, a Sino-Tibetan Kiranti language of Nepal, provides a particularly compelling example of this sort. Verbal prefixes in Chintang can be freely ordered, as shown in (4). The result of this morphotactic variability is that certain morphosyntactic feature sets can be realized in up to six different ways (Bickel et al., 2007).

(4) Free prefix ordering in Chintang[1]

a.	u-	kha-	ma-	cop	-yokt	-e
	3NS.A-	1NS.P-	NEG-	see	-NEG	-PST
b.	u-	ma-	kha-	cop	-yokt	-e
	3NS.A-	NEG-	1NS.P-	see	-NEG	-PST
c.	kha-	u-	ma-	cop	-yokt	-e
	1NS.P-	3NS.A-	NEG-	see	-NEG	-PST
d.	ma-	u-	kha-	cop	-yokt	-e
	NEG-	3NS.A-	1NS.P-	see	-NEG	-PST
e.	kha-	ma-	u-	cop	-yokt	-e
	1NS.P-	NEG-	3NS.A-	see	-NEG	-PST
f.	ma-	kha-	u-	cop	-yokt	-e
	NEG-	1NS.P-	3NS.A-	see	-NEG	-PST

All: 'They didn't see us.'

The examples in (4) all contain a verbal stem meaning SEE that is inflected for negative polarity and past tense, resulting in the form -*copyokte*. This complex is preceded by three different prefixes: a nonsingular third-person agent marker (*u*-), a nonsingular first-person patient marker (*kha*-), and a negation marker (*ma*-). These prefixes can be freely ordered such that all logical variations in sequence are possible, with no attendant change in propositional content. As it turns out, this kind of optionality is not particularly rare, and is attested in the inflectional and derivational morphologies of a number of other languages. Similar phenomena, for example, have been noted in Maari (Luutonen, 1997), Nhanda (J. Blevins, 2001), Kusunda (Watters, 2006), Madurese (Stevens, 1971), and Totonac (McFarland, 2007).

[1] NS = nonsingular, 1 = first person, 3 = third person, A = actor, P = primary object, NEG = negative, PST = past.

Bulgarian provides another example of this kind of remarkable morphological variability. Among Slavic languages, Bulgarian boasts a particularly rich system of verbal inflection. Bulgarian verbal predicates can be marked for person, number, tense, aspect, mood, voice, evidentiality and polarity. In many cases, these predicates are realized as multi-word forms (i.e., periphrases) that undoubtedly have syntactic origins, but are best analyzed—based on their decidedly non-compositional nature (Ackerman & Stump, 2004)—as belonging to morphological paradigms (Manova, 2007). An example is given in (5).

(5) The Bulgarian negative future perfect (Manova, 2006, 2007)[2]

a.	ne	šte	săm	pisal
	NEG	FUT	be.1.SG.PRES	write.AOR
b.	njama	da	săm	pisal
	not have.PRES	COMPL	be.1.SG.PRES	write.AOR
c.	ne	šte	băda	pisal
	NEG	FUT	be.1.SG.PRES	write.AOR
d.	njama	da	băda	pisal
	not have.PRES	COMPL	be.1.SG.PRES	write.AOR

All: ‘I will not have written.’

The above predicates are specified for the future tense, despite the fact that parts of each expression (i.e., the auxiliaries *săm* and *băda*, and the negation marker *njama*) are normally associated with the present tense. It is thus not possible to treat (5a-d) as syntactically and semantically compositional in an incremental fashion—such treatment would lead to the uninterpretable unification of future and present tense. Instead,

[2] NEG = negative, COMPL = complementizer, FUT = future, PRES = present, 1 = first person, SG = singular, AOR = aorist participle

speakers are only able to arrive at the correct interpretation by considering the entire periphrase to be semantically atomic. This entails that meaning can be lexically assigned over multi-word forms, a hypothesis that finds a great deal of support in recent research within the revived tradition of Word and Paradigm morphology (Ackerman & LeSourd, 1997; Ackerman & Malouf, 2007; Ackerman & Stump, 2004; Ackerman & Webelhuth, 1998; Spencer, 2004; Stewart & Stump, 2007; Stump, 2002).[3] Under this kind of analysis, each expression in (5) is morphologically generated, with all four expressions simply representing different surface realizations of the same lexical-semantic and grammatical content. Similar instances of optionality are available in a number of other Bulgarian verbal paradigms (Manova, 2006). These data from Bulgarian—along with the previous example from Chintang—thus serve to illustrate some of the more unconventional and inventive ways in which optionality is attested in morphologically rich languages.

Another common type of morphological optionality—one that is somewhat more modest in nature than the above examples but still raises many theoretical questions—commonly occurs in the case of synthetic-analytic alternations. The literature is replete with examples showing that, within the same paradigm, some cells can be lexically realized using one-word *synthetic* expressions, while others are lexically realized using multi-word *analytic* (or alternatively, *periphrastic*) expressions.[4] This is true, for example, of verbal predicates in Mari and Udmurt (Ackerman & Stump, 2004), the Latin perfective passive (Sadler & Spencer, 2000), Irish conditionals (Andrews, 1990) and

---

[3] Chapter 5 provides a detailed discussion of Word and Paradigm models of morphology.
[4] The terms *analytic* and *periphrastic* will be used interchangeably throughout this thesis to constrast multi-word expressions with one-word, *synthetic* expressions.

nominal number and class in Malto (Stump, 2002). In some instances, the synthetic and analytic patterns of inflection can both be applied to the same cell so that one-word and multi-word expression realize identical semantic content. This is the case for some English comparative adjectives (Aronoff, 1976; Graziano-King, 2003; Mondorf, 2003), and for Greek plurals inflected for the genitive, as in the example below (Sims, 2005).

(6) Synthetic-analytic optionality in the Greek plural genitive[5]

a. μια παρέα εννέα γυναικ-ών
mia paréa innéa γinek-ón
a group nine women-GEN.PL

b. μια παρέα με εννέα γυναίκ-ες
mia paréa me ennéa γinék-es
a group with nine women-GEN.PL

Both: 'a group of nine women'

In (6), the women (*γυναικ*) are conceived of as belonging to a group (*παρέα*). This idea can be signaled synthetically through the use of the genitive plural suffix -*ών* on *γυναικ*, as in (6a). The same notion, however, can also be realized analytically in a prepositional phrase in which γυναικ occurs as a noun following the prepositional head *με* ('with'). Both expressions are associated with the same interpretation.

The preceding examples and references illustrate the pervasive nature of morphological optionality in natural language. The use in morphology of multiple expressions to signify the same content is, accordingly, a phenomenon that must be addressed in all theoretical frameworks. Moreover, the fact that optionality occurs in productive (inflectional) morphology and is not simply limited to so-called irregular formations makes it a phenomenon that demands exploration and analysis. All of these

---

5 GEN = genitive, PL = plural, ACC = accusative.

factors motivate the need for experimental studies of morphological optionality on par with those already available on optionality in the syntactic domain. In the remainder of this chapter, I provide an initial foray into this area by examining morphological optionality as it occurs among the relatively large class of English adjectives that can be inflected for the comparative either synthetically, or analytically (Aronoff, 1976; Frank, 1972; Graziano-King, 2003), along the lines illustrated for Greek in (6).

1.3 English Comparative Inflection

English adjectives can be inflected for comparison synthetically via *-er* suffixation (e.g., *bigger*, *older*, *easier*), or analytically through combination with the independent element *more* (*more interesting*, *more beautiful*, *more comfortable*). A few adjectives form the comparative suppletively (*worse*, *better*, *further*), although it should be observed that some of these consist of a suppletive stem (e.g., *bett-*, *furth-*) that is still combined with the synthetic *-er* suffix. Originally, all English adjectives were inflected synthetically, as is still generally the case across Germanic. The first attested examples of analytic comparison come from Old English (OE), and occurred in the latter half of the ninth century (Díaz, 2006). Widespread use of the analytic, however, did not become common until well into the Middle English period, and it was not until the onset of early Modern English, circa 1650, that analytic comparatives could be said to occur as frequently as they do today (Kytö & Romaine, 1997, 2000).

There is some debate over the origin of the analytic pattern of inflection. Some authors have argued that it was borrowed into English under influence from French and Latin (Kytö & Romaine, 1997, 2000). The fact, however, that comparatives were being formed analytically in English prior to the Norman conquest in 1066—the first occasion on which large numbers of French and English speakers came into sustained contact—argues in favor of an endogenous genesis. Díaz (2006) claims that *more* originally functioned in OE as a verbal modifier, and that its emerging use as an adjectival degree modifier came about through combination with participles, whose lexical category status can often be ambiguous. For example, in the OE sentence *Ianus wæs ma gelæred* ('Janus was more learned'), the participle *gelæred* can be viewed as either a verb in a passive-like construction, or an adjective in a copula construction. In the former case, *ma* ('more') is an adverb; in the latter, it acts as an adjectival degree modifier in an analytic comparative construction. According to Díaz, the use of *ma* as an adjectival modifier was extended by analogy over time from these adjective-like participles to the class of prototypical adjectives.

Regardless of the etiology of the analytic, in the current synchronic grammar the system is structured in large part around a correlation between inflectional class and stem length (Aronoff, 1976; Embick, 2007; Frank, 1972; Poser, 1992; Stump, 2001). Most monosyllabic stems, for example, follow the synthetic pattern of inflection, while most stems of two or more syllables are inflected analytically. This generalization is subject to a number of exceptions, however. Disyllabic stems with certain types of end-segments (e.g., *-y*, *-ow*, *-some*, and *-er*) often have both options available (*angrier* ~ *more angry*, *mellower* ~ *more mellow*, *handsomer* ~ *more handsome*; Aronoff, 1976; Frank, 1972), and monosyllabic stems do not uniformly participate in the synthetic pattern. Synthesis appears to be blocked, for instance, in participles like *bent* and *known* (e.g., **benter*, **knowner*). Such forms have hybrid distributions that set them apart from prototypical adjectives (Díaz, 2006). The synthetic pattern is similarly dispreferred among low frequency, non-gradable monosyllabics. Graziano-King (2003) collected ratings and forced-choice judgment data over a range of comparative forms, and found that synthesis was the inflectional method of choice for three classes of stems: high-frequency gradable monosyllabics, disyllabics ending in -*y* regardless of frequency, and disyllabics with other ending types (e.g., -*le*, -*ow*, -*some*, and -*er*) as long as their frequency was at least 61 per million tokens in the Francis and Kucera (1982) corpus. This result explains why stems like *apt* and *just* tend to robustly eschew synthesis (**apter*, **juster*, cf. *more apt* and *more just*): both are low frequency, and *just* is decidedly non-gradable.

Distributional observations along these lines all tend to point towards a conclusion that is reflected in a wide range of linguistic analyses (Embick, 2007; Poser, 1992; Stump, 2001): the analytic pattern has default status in the grammar. Whereas the analytic has no constraints on its applicability, the synthetic appears to be restricted on the basis of syllabicity, stem end-segment, frequency, gradability, and syntactic class. Note, however, that these attributes may account for the relative stability of the synthetic in the face of an imperialistic class of *more* comparatives. Computational models of acquisition demonstrate that smaller inflectional classes—like the synthetic—can be maintained over the course of ontogeny under certain conditions. These less productive classes are resistant to takeover when they have high token frequencies, and also when they can be partitioned away from larger classes (like the analytic) based on shared features (Plunkett & Marchman, 1991, 1993). As the preceding discussion indicates, adjectives that inflect synthetically are generally higher in frequency, and tend to cluster together along a number of dimensions. Their shared phonological and semantic characteristics may give them a learnability advantage that explains why the current division of adjectives into analytic and synthetic categories—in which analytics outnumber synthetics by roughly four to one (Boyd, 2006)—has persisted over the past 400 years.

1.4 Comparative Optionality

The focus of the present work is on the class of adjectives that participate in both the synthetic and the analytic patterns of inflection. Here, I discuss the results of a brief corpus survey of these items, with the goal of outlining some of their shared features and tendencies. To begin, I identified all stems in the 10-million-word spoken section of the British National Corpus (BNC; 2001) that showed some variation in comparative form (i.e., the set of adjectives that were inflected at least once in the corpus with -*er*, and at least once with *more*). This amounts to 402 items. For each of these, I then calculated their probability of synthetic inflection. The figure below shows all 402 adjectives, situated in two-dimensional space according to the log frequency with which they appeared in the comparative (in either form), and their probability of synthetic inflection—written as P(*er*).

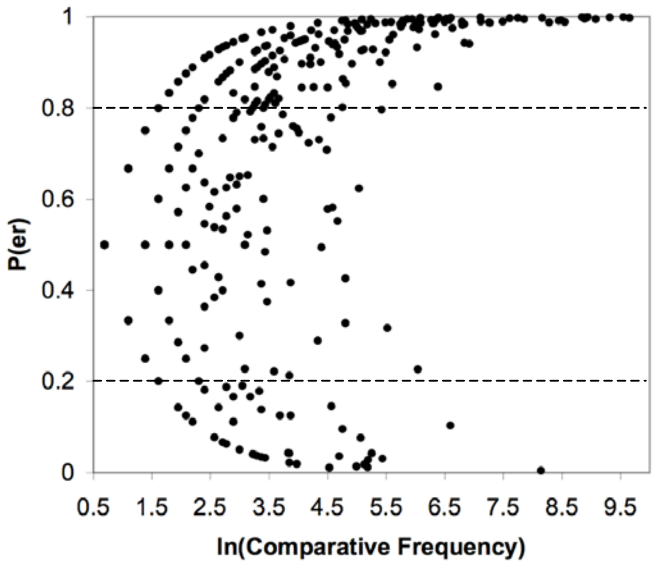


Figure 1.1: Distribution of adjectives by comparative frequency and probability of *-er* inflection.

All of the adjectives plotted above show at least some variation in the way in which they inflect for the comparative. As P(*er*) approaches one or zero, however, this variability diminishes such that adjectives near the top of the figure are nearly always inflected with *-er*, and adjectives near the bottom are nearly always inflected with *more*. Based on the C-shape of the plot, there appears to be a systematic relationship between frequency and P(*er*): the most variability in comparative form occurs among lower frequency adjectives. The implications of this tendency will be addressed shortly.

I assessed whether certain kinds of phonological shapes might be more amenable to optional inflection than others by first restricting the sample of 402 stems to only those items that showed somewhat larger amounts of variability—i.e., those that did not strongly favor one inflectional pattern or the other. This was accomplished by eliminating all adjectives from the sample that had a probability of synthetic inflection greater than 0.8, or less than 0.2. This procedure leaves 185 stems, which are located between the two horizontal dashed lines in Figure 1. Of these, a large majority (68%) are disyllabic, while only 29% and 3%, respectively, are monosyllabic and trisyllabic. An examination of the kinds of endings available for these items indicates that monosyllabic stems that participate in optionality do not cluster according to end-segment. Disyllabics, however, do tend to group together. A full 59% of the 185 stems examined are disyllables in *-y*. A variety of other ending types are additionally available among disyllabics, including the *-le*, *-ow*, *-some*, and *-er* patterns mentioned in the work of Aronoff (1976), Frank (1972), and Graziano-King (2003). None of these, however, account for more than 4% of the 185 stems under consideration. In sum, the prototypical optional adjective appears to be a low frequency disyllabic in *-y*, e.g. *hazy*, *salty*, and *roomy*.

The frequency and phonological characteristics of adjectives that participate in optionality indicate that speakers must deal with conflicting cues when deciding how to inflect these forms. Optional adjectives tend to have phonological shapes that are known to be accommodating to *-er* suffixation, e.g. monosyllabics, or disyllabics in -y, *-le*, *-ow*, *-some*, and *-er*. As a result, a simple consideration of the phonological cues predicts that, all else being equal, they should be inflected synthetically. The frequency cues, however, go in the opposite direction. Synthesis is generally only associated with high frequency stems (Graziano-King, 2003), so the low frequency of adjectives that participate in optionality militates against synthesis and in favor of use of the analytic. Conflicting phonological and frequency factors thus conspire to create a domain in which speakers are unsure exactly which inflectional pattern to use. The result is optionality.

1.5 Why study comparative optionality?

As previously mentioned, the main goal of the current research is to provide an initial foray into the psycholinguistic and theoretical study of comparative optionality. Syntactic optionality has been investigated in extensive detail using experimental, computational, and corpus methods (Bresnan, 2006; Bresnan et al., 2007; Chang et al., 2006; V. S. Ferreira, 1996; V. S. Ferreira & Dell, 2000; V. S. Ferreira & Firato, 2002; Jaeger, 2006; Roland et al., 2006; Savage et al., 2003, 2006; Slevc, 2007; Temperley, 2003). This line of research has been enormously profitable in terms of generating new insights into the nature of language processing and learning. Its very success, in fact, highlights the potential benefits that are likely to accrue when the same kinds of investigatory techniques are applied to the study of morphological optionality.

There are already hints of the viability of this research direction, and its ability to inform our understanding of language. For example, a series of corpus analyses by Mondorf (2002, 2003) indicate that comparative inflection is driven by multiple weighted factors. This outcome is in line with recent findings from the study of other grammatical alternations (Bresnan, 2006; Bresnan et al., 2007; Jaeger, 2006; Roland et al., 2006) that suggest that language production is probabilistic—a proposition that goes against certain strongly-held generativist assumptions (Aarts, 2004). Mondorf's work is additionally of interest because of what it may say about the functional benefits of morphological optionality. A number of researchers have collected data demonstrating that the flexibility afforded by syntactic optionality allows speakers to strategically produce those parts of a sentence that they are ready to articulate, and to delay those parts that require more processing (V. S. Ferreira, 1996; V. S. Ferreira & Dell, 2000; V. S. Ferreira &

Firato, 2002; Slevc, 2007). The choice between competing comparative options may confer the same kind of advantage by allowing a speaker to speed up or delay the production of information downstream from the comparative. Mondorf (2003) found that, for adjectives that participated in comparative optionality, processing load was positively correlated with use of the analytic pattern of inflection: speakers tended to use the durationally longer *more* variant more often when under higher loads. This behavior is consistent with the use of the analytic as a stalling tactic. As downstream material becomes more difficult to process, speakers appear to be able to use *more* to buy additional processing time.

The study of comparative optionality is also of significant theoretical interest. A number of researchers have used the comparative system as an example of what are known as *blocking effects* in morphology (Embick, 2007; Poser, 1992; Stump, 2001). Chapter 5 describes this phenomenon and the analyses that have been offered of it in detail. For the present, it is sufficient to note that blocking is synonymous with the *lack* of optionality. The prevalence of blocking analyses of the comparative thus seems to reflect the widespread assumption among linguistic theorists that optionality is nonexistent in this domain, or is at least so circumscribed as to be ignorable. Some of these same analyses are additionally used to make unsupported claims about the division of morphological and syntactic processes in grammar (Poser, 1992), and even to argue against morphology as an independent linguistic domain (Embick, 2007). Overall, there is a tendency in the theoretical literature to make erroneous assertions about what the data look like (i.e., that there is no optionality), and then to use this as support for dubious claims about the nature of grammar. Given these circumstances it becomes increasingly important to offer clear evidence not only of the existence of comparative optionality, but also of its systematic functional importance to speakers of English.

1.6 Organization of the Following Chapters

The chapters that follow provide a detailed study of English comparative optionality that includes results from comprehension, production and acquisition.

In Chapter 2, I investigate a hypothesis proposed in Mondorf (2003) that, among adjectives that can be inflected both analytically and synthetically, listeners find the analytic variant to be simpler to process. I use offline (Experiment 1) and online (Experiment 2) measures to test this notion, and follow up in Experiment 3 with a ratings study designed to determine why analytics might be preferred.

Chapter 3 describes an elicited production experiment (Experiment 4) designed to test Mondorf's (2003) claim that speakers increase their use of the analytic in complex processing environments. It additionally evaluates the possibility that such an increase might be part of an *audience design* strategy geared towards assisting listener processing (Haywood et al., 2005; Keysar, Barr, & Horton, 1998; Kraljic & Brennan, 2005).

In Chapter 4, I turn to the acquisition of comparative inflection. Past studies in this area have shown that comparative development is characterized by a single stage of overgeneralization, in which children erroneously apply the *-er* suffix to stems that adults inflect with *more* (e.g., **dangerouser*, **beautifuler*). What seems odd about this finding is that no researcher has ever documented a stage of *more* overgeneralization (e.g., **more fast*, **more easy*, cf. adults' *faster, easier*). Such a stage is predicted based on widespread agreement that the analytic pattern is the grammatical default in English

(Embick, 2007; Poser, 1992; Stump, 2001). I hypothesize that the lack of evidence in favor of a stage of *more* overgeneralization is due to a design flaw in the elicitation method used by previous researchers, and offer a new method—developed in Experiments 5 and 6—that promises to resolve this issue, as well as provide insight into the developmental behavior of adjectives that participate in optionality.

I conclude in Chapter 5 by discussing how the experimental findings of Chapters 2, 3 and 4 bear on theory construction. The dominant trend among grammatical theorists has been—as mentioned in §1.5—to treat the comparative system as an example par excellence of morphological blocking, where each lexeme (or word) is categorically associated with the possibility of either synthetic or analytic expression (Embick, 2007; Poser, 1992; Stump, 2001). I demonstrate that categorical blocking effects, when they occur, are best analyzed as highly probable mappings subsumable within a stochastic approach that treats them and the optional selection typified by English comparatives in a uniform fashion. This more empirically adequate theory of comparative inflection is argued to be straightforwardly formulable within a Word and Paradigm model of morphology (Bochner, 1993).

2. Comprehension

Mondorf (2003) studied the effects of complexity on English comparative inflection in a 914 million-word corpus. Complexity was defined in multiple ways—syntactically, morphologically, phonologically and semantically—but the study's main claims concern complexity as indexed by the presence or absence of an infinitival complement following the comparative. Invented examples of sentences that vary according to this feature are given below.

(1) Complexity in Mondorf (2003)

a. Complex: Kenny G is more mellow to listen to than Yanni.

b. Simple: Kenny G is mellower than Yanni.

The sentence in (1a) contains an infinitival complement (underlined) following the comparative. Mondorf coded sentences of this sort as *complex*. Sentences, however, that contained a comparative without an infinitival complement—like (1b)—were coded as *simple*. Each sentence was also coded for morphology type: whether the comparative was realized synthetically with *-er*, or analytically with *more*. This design allowed Mondorf to treat complexity as an independent variable, and to use it to investigate the effect of complexity on the rate of analytic inflection for 28 adjectives that participated in both inflectional patterns. Her findings are summarized in Figure 2.1, and indicate that the rate of analytic use is higher when the comparative occurs in complex environments (i.e., preceding an infinitival complement) than when it occurs in simple environments.

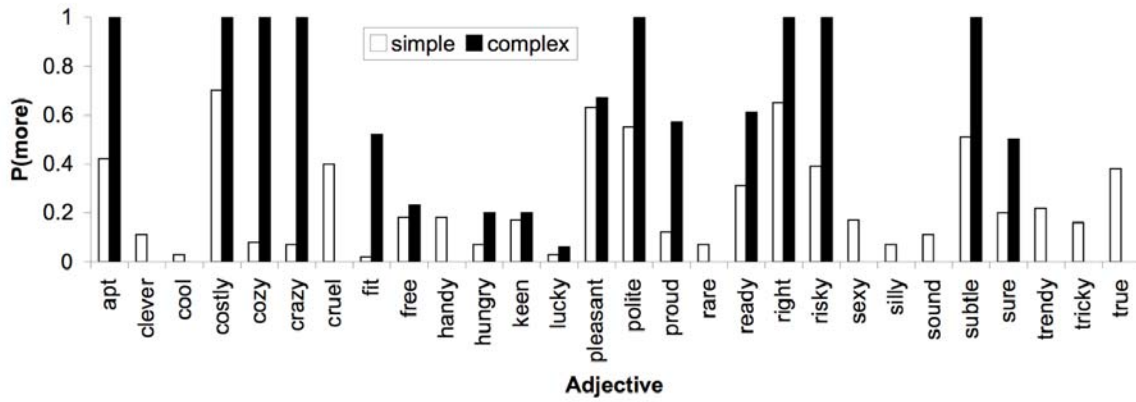


Figure 2.1: Effects of complexity on rate of analytic usage (Mondorf, 2003).

Note that some of the adjectives shown above do not have black bars. This is because Mondorf's (2003) corpus did not contain sentences in which these items appeared in comparative form followed by an infinitival complement. For the 17 adjectives that did have tokens available in both conditions however, a reliable effect of complexity appears to be visible: the difference in rate of analytic use always goes in the predicted direction, with the analytic being increasingly preferred as a function of complexity.

Mondorf (2003) interprets this effect as an *audience design* measure (V. S. Ferreira & Dell, 2000; Haywood et al., 2005; Keysar et al., 1998; Kraljic & Brennan, 2005; Temperley, 2003). Audience design refers to the hypothesis that speakers are sensitive to the needs of their addressees, and formulate their utterances to facilitate addressee processing.[6] According to Mondorf, this plays out in the comparative data in the following way. The addition of an infinitival complement leads to an increased

[6] For purposes of expositional ease, I use the term *speakers* throughout to refer to both speakers and writers, and *listeners* (or *addressees*) to refer to listeners and readers. While there are undoubtedly modality-specific effects in both production and comprehension, this issue is not addressed in the present work.

processing load. Speakers attempt to compensate for the effects of an increased load on their listeners by choosing to use the comparative variant that is simplest for listeners to process—i.e., the analytic. While this interpretation is consistent with the findings in Figure 2.1, it begs empirical verification. I address the audience design claim in detail in Chapter 3. The current chapter however, is devoted to investigating the hypothesis that the analytic is actually easier for listeners to process. Mondorf's (2003) findings are based on corpus data and are consequently unable to speak directly to questions related to comprehension. This motivates the use of comprehension-specific methods to determine *whether* the analytic confers a processing advantage on listeners, and *why* this might be the case.

2.1 Experiment 1

I refer to the notion that analytic comparatives are easier for addressees to process than synthetic comparatives as the *analytic advantage hypothesis* (AAH). Experiment 1 offers a simple test of the AAH. I methodically manipulate morphology type (analytic versus synthetic) in a series of stimulus sentences, and measure comprehension preferences by condition.

(2) Manipulating morphology type

a. Analytic: Kenny G is more mellow to listen to than Yanni.

b. Synthetic: Kenny G is mellower to listen to than Yanni.

Both (2a) and (2b) are complex in that they contain a comparative followed by an infinitival complement. The only difference between the two is that the analytic variant is used in (2a), and the synthetic variant is used in (2b). If morphology type has an effect on comprehension in the way that the AAH predicts, then we would expect that participants would prefer sentences like (2a) over sentences like (2b) on a variety of comprehension measures. In Experiment 1, I use an offline measure—acceptability ratings—to assess whether or not this expectation is warranted.

Experiment 1 focuses on complex sentences and not simple sentences in an attempt to give the AAH the best possible chance of succeeding. According to Mondorf's (2003) interpretation of the data pattern in Figure 2.1, speaker use of the analytic is particularly helpful in complex processing environments, where processing loads on addressees are high. This account suggests that use of the analytic might not be equally helpful in simple environments because addressees are working under low loads and are able to devote more resources to comparative processing. That is, a floor effect that obscures the analytic advantage may occur under low loads. This possibility motivates an initial focus on processing in complex environments.

2.1.1 Participants

95 UCSD undergraduates from an introductory course in language acquisition participated for course credit.

2.1.2 Materials

Stimulus sentences were created using 18 different adjectives. Seven of these (*apt*, *fit*, *free*, *keen*, *proud*, *ready* and *pleasant*) were identified in Mondorf (2003) as items that showed increased use of the analytic in the complex condition, and occurred in the corpus enough times that confidence in the P(*more*) estimates given in Figure 2.1 was high. Another five (*crazy*, *hungry*, *lucky*, *risky* and *sure*) were argued to provide weaker support based on the notion that they were also favored *more* in complex environments, but were represented by too few tokens in the corpus to know whether this trend was reliable. This group of 12 was supplemented with an additional six adjectives (*angry*, *cruel*, *lively*, *mellow*, *sorry* and *stark*), which were selected based on counts from the BNC (2001). We identified all adjectives that varied in comparative form in the 10 million-word spoken section of the BNC, then chose six that appeared in the comparative at least 20 times (the sum of the *-er* and *more* tokens for that adjective), and had roughly equal *-er* and *more* counts. Use of this method served to increase the set of test adjectives available to us by identifying items that should be amenable to either *-er* suffixation, or inflection with *more*.

36 *critical sentences* were constructed using the 18 adjectives given above. Each critical sentence followed the pattern in (2), and contained an adjective in comparative form immediately followed by an infinitival complement. For each adjective, two sentences were created. These constituted a minimal pair that differed only in terms of how the comparative was realized: either analytically (as in 2a), or synthetically (as in 2b). Additionally, 18 *filler sentences* were constructed. Filler sentences did not contain comparative adjectives. They functioned as washout trials whose purpose was to allow the representations activated when dealing with a critical sentence to return to baseline before the onset of the next critical sentence (Anderson, 2001). All critical sentences are listed in Appendix A.

Paper-and-pencil ratings questionnaires were created using different combinations of the 36 test sentences and 18 filler sentences. Specifically, each questionnaire was constructed by printing one randomly-selected filler sentence, then following it with a randomly-selected member of a critical pair (e.g., either 2a or 2b, above). Once one member of a test pair had been selected, its counterpart was made unavailable for selection. This process was iterated until the questionnaire contained 36 items. Use of this method ensured that test and filler sentences alternated, that each participant provided ratings for 18 test sentences and 18 filler sentences, and that each participant saw only one member of a critical sentence pair. 95 questionnaires were constructed in this fashion altogether—one for each participant. A 7-point ratings scale was printed immediately beneath each questionnaire item. Each scale had the words “Completely normal” printed under the numbers at the high end of the scale—6 and 7. “Okay” was printed under 3, 4 and 5, and “Completely bizarre” was printed under 1 and 2. The first page of each questionnaire consisted of questions about participants’ language backgrounds.

2.1.3 Procedure

Each participant received a questionnaire to fill out. Participants were told to complete the questionnaire on their own time and to return it within a week. Instructions printed at the beginning of each questionnaire asked participants to individually rate the acceptability of all questionnaire sentences by circling a number on the 7-point scale printed beneath each sentence.

2.1.4 Results

Access to the student subjects used in Experiment 1 was contingent on allowing any student who wanted to participate the opportunity to do so. This resulted in 19 non-native speakers of English being tested. These participants were excluded from all subsequent analyses, leaving a total of 76 native speakers. This group participated in 1,368 total critical trials. Two subjects, however, failed to rate one sentence apiece, so the analyses below are based on data from 1,366 trials. Overall, participants seemed to feel that most of the critical sentences they considered were quite native-like. High ratings—a six or seven—were given on 46.3% of all trials. A similar proportion of trials (42.3%) were given medium ratings—threes, fours, or fives. Only 11.4% of trials received a low rating—either a one or a two.

When considering ratings behavior by morphology type, we found that critical sentences in *more* were rated more than a point higher than sentences in *-er* (5.46 versus 4.39). To ascertain whether this difference was reliable, we conducted separate ANOVAs using morphology type (*-er* versus *more*) as a repeated-measures variable, and modeling subjects (F_1) and items (F_2) as random effects. These analyses show that the difference in ratings between sentences containing analytic comparatives and those containing synthetic comparatives is significant by subjects, $F_1(1, 75) = 73.80$, $p < .0001$, and by items, $F_2(1, 17) = 13.30$, $p < .002$. Results are depicted graphically below.

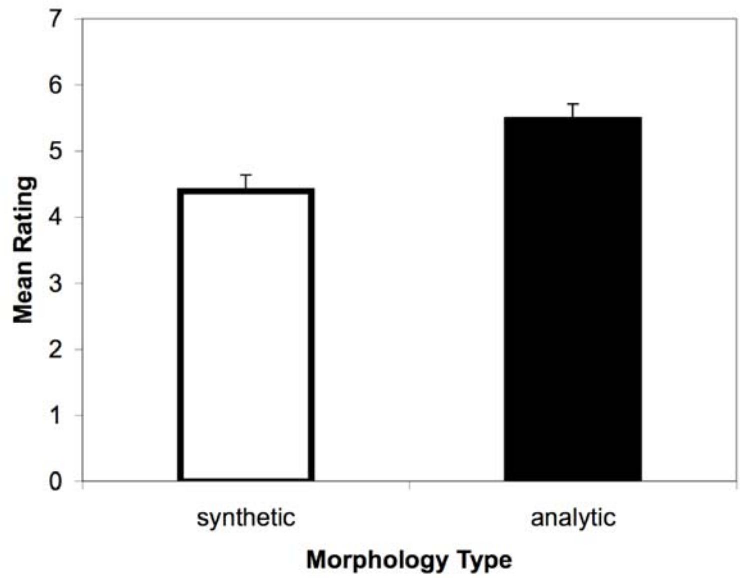


Figure 2.2: Effects of morphology type on sentence acceptability.

That participants in Experiment 1 found sentences in *more* to be significantly more acceptable than the same sentences in *-er* suggests that Mondorf (2003) was correct in claiming that analytic comparatives are simpler for listeners to process. This result would appear to go in favor of the AAH.

Acceptability ratings, however, may not be the most valid measure of processing ease or difficulty. In particular, because Experiment 1's participants completed their ratings questionnaires on their own time, over the course of a week, it is impossible to know whether non-processing-related factors may have influenced how the ratings task was completed. Given enough time to consider an item, all sorts of factors may have been brought to bear in deciding how to rate it—stylistic considerations, explicit comparison to other items, etc. This motivates a replication of Experiment 1 using a design in which the relationship between processing and the preference for analytic morphology is made much more direct.

2.2 Experiment 2

The goal of Experiment 2 was to verify the findings of Experiment 1 using an online measure of processing difficulty. This is motivated by the potential confounds that arise when attempting to assess processing difficulty using an offline measure such as a ratings task. Because the offline task is performed over a relatively slow time scale—particularly in this case, where participants were given a full week to mull over their responses—it is difficult to know whether the results accurately reflect the kinds of processing that are associated with language comprehension as it occurs in real-world situations. To remedy this problem, Experiment 2 supplements the findings of Experiment 1 with data from a self-paced reading task that more closely mimics the millisecond-by-millisecond processing difficulties associated with comprehension.

2.2.1 Participants

Twenty native speakers of English were recruited from a UCSD undergraduate course on language in politics and advertising. All participants were compensated with course credit.

2.2.2 Materials

Four lists were constructed using the same 36 critical sentences from Experiment 1, and a new set of 18 filler sentences. This new set of fillers was more closely matched to the critical sentences in length than the fillers of Experiment 1.

Each list was 72 items long, and was divided into two blocks—Block 1 and Block 2. Each block contained 36 sentences; half of these were filler sentences, and the other half were critical sentences. Filler sentences were randomly distributed in Block 1 of

Lists 1 and 2 at odd-numbered positions. Critical sentences were pseudo-randomly distributed in Block 1 of Lists 1 and 2 at even-numbered positions with the restriction that each block contained 9 critical sentence in *-er*, and 9 critical sentences in *more*, and that no adjective was represented more than once. In Block 1 of Lists 3 and 4, the position preferences were switched—fillers were distributed in odd-numbered positions, and critical sentences were distributed in even-numbered positions. For each of the four lists, Block 2 was constructed by flipping the morphology type of the comparatives in the Block 1 critical sentences. For example, if Block 1 contained a sentence with *more pleasant* in it at position 6, then the same sentence occurred at position 6 in Block 2, but with *pleasanter*. The critical sentences in Block 2 of each list were thus the complement of the critical sentences in Block 1 in terms of how the comparative was realized. The filler sentences in Block 2 of each list were repeated verbatim from Block 1, in the same order and positions. The order of appearance of each specific critical sentence was additionally counterbalanced across lists. This means, for instance, that the critical sentence with *more proud* occurred in Block 1 on two lists, and in Block 2 on the others.

2.2.3 Procedure

Participants were tested one at a time in a quiet room, and were randomly assigned to read one of the four lists described above. Each participant was seated in front of a computer monitor, with a standard Psyscope button box within easy reaching distance. The button box's green button was labeled *true*, and the red button was labeled *false*. The central yellow button was unlabeled.

Participants were handed a sheet of paper with written instructions. The instructions told them that they would be reading sentences one word at a time on the monitor, and that they could advance through each sentence by pressing the yellow button. After each sentence was read, they were told that they would be answering a true-false comprehension question by pressing the appropriate button on the button box. The instructions encouraged participants to read for speed and accuracy.

Each trial began with a fixation cross displayed at the center of the computer monitor. To begin reading, participants pressed the yellow button. This displayed the first word of the current sentence at the center of the monitor. To advance through a sentence, participants simply pressed the yellow button again for each word. Pressing the yellow button recorded a reading time for the currently displayed word, and advanced to the next word in the sentence. The end of each sentence was labeled with a period. After finishing a sentence, participants were presented with a true-false comprehension question, which could be answered by pressing either the green (true), or red (false) button box button. These questions were designed to be easily answerable if the participant had been paying attention.

Each participant began by reading through 10 training sentences in order to familiarize themselves with the procedure. After that, Block 1 of the list that they had been assigned to was read. Upon finishing Block 1, participants were allowed a short break. Before beginning Block 2, participants were told that they would be reading the same sentences that they had read in the first half of the experiment so that we could verify the accuracy of our measurements. Based on debriefing interviews conducted with each participant after the experiment had concluded, we determined that no participant had noticed that the sentences in Blocks 1 and 2 were subtly different from one another.

An experimenter was present during all phases of testing in case participants had any questions. The experiment was controlled using Psyscope (Cohen, MacWhinney, Flatt, & Provost, 1993).

2.2.4 Results

If the AAH is correct, and analytic comparatives are simpler to process than synthetic comparatives, then all else being equal, analytics should be read faster than synthetics. The current design, however, in which words are presented one screen at a time, makes it difficult to compare analytic and synthetic reading times directly. This is because the two-word analytic is presented on two separate screens. As a means of circumventing this problem, analyses were conducted on the infinitival maker *to*, which occurred after the comparative in each of the critical sentences. If the analytic does confer some sort of processing advantage, then *to*'s following analytic comparatives should be read faster than *to*'s following synthetic comparatives.

Since reaction time distributions are highly non-normal, I log-transformed the reading times for *to* in each of the 720 critical trials that were run. The data were then trimmed on a subject-by-subject basis by removing all datapoints greater than two standard deviations from the mean. This process resulted in the exclusion of 16 outliers. Finally, all trials in which participants failed to correctly answer the comprehension question were excluded. This amounted to another 35 trials, and left us with analyzable reading times for 669 trials. These were divided roughly equally between trials in which the comparative was realized analytically (340 measurements), and trials in which it was realized synthetically (329 measurements).

For this set of datapoints, *to* was read in an average of 289 milliseconds. When considered by morphology type, *to* was read faster in critical sentences containing analytic comparatives (281 ms) than in the same sentences with synthetic comparatives (297 ms). To ascertain whether this difference was reliable, I conducted separate ANOVAs on the log-transformed data using morphology type (*-er* versus *more*) as a repeated-measures variable, and modeling subjects (F_1) and items (F_2) as random effects. These analyses demonstrate a significant main effect of morphology type by subjects, $F_1(1, 19) = 25.23$, $p < .0001$, and by items, $F_2(1, 17) = 6.30$, $p < .03$. This outcome is illustrated below.

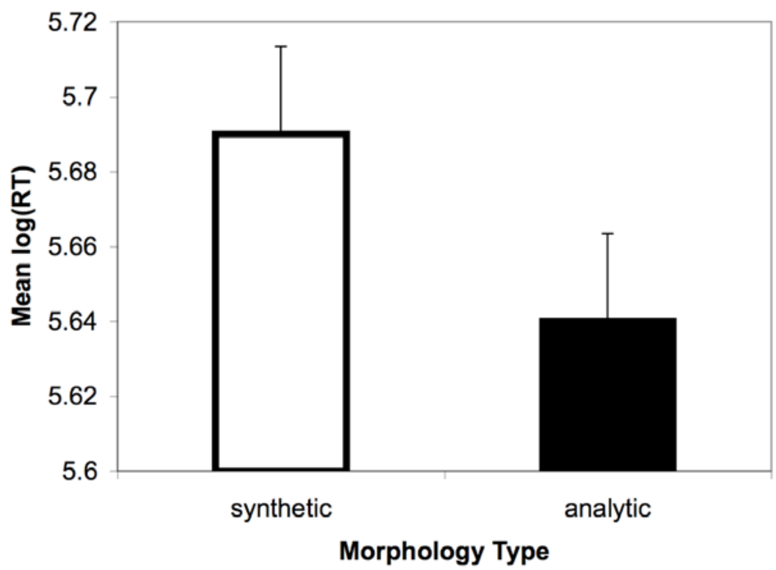


Figure 2.3: Effects of morphology type on the processing of infinitival *to*

Because sentences in the analytic and synthetic conditions differed solely according to morphology type, I conclude that the significant reading time differences for infinitival *to* are the result of a spillover effect. The processing advantage afforded by the analytic (or, alternatively, the penalty associated with the analytic) was large enough to affect the reading times of words downstream from the comparative.

These findings replicate the comprehension preference for the analytic found in Experiment 1, and do so using an online methodology that avoids some of the problems associated with a protracted weeklong ratings procedure. However, while the results of Experiments 1 and 2 are consistent with the AAH, neither experiment is able to say anything about *why* analytic comparatives might be favored in comprehension. In Experiment 3, I attempt to answer this question by testing two different possibilities in a new ratings study.

2.3 Experiment 3

Mondorf (2003) suggests that analytic comparatives might be preferred in complex environments for two reasons. According to what I will call the *warning hypothesis* (WH), the analytic acts as a conventionalized signal to addressees of upcoming complexity. Speakers look ahead in their own processing, and if they determine that downstream material may be potentially difficult for addressees to handle, they choose to produce the analytic comparative as a warning. Addressees are consequently able to prepare by strategically allocating processing resources in a way that helps to mitigate the negative effects of complexity. An alternative to the WH is the *parsing hypothesis* (PH). According to the PH, the analytic is preferred in complex environments because it allows for earlier phrase structure recognition, and avoids a temporary parsing ambiguity that is associated with the synthetic.

As an example of this ambiguity, consider the string *The senator was angr*[i] ______. Possible continuations include (i) *and tired*, and (ii) *-er than usual yesterday*. In (i), *angry* is parsed as a bare stem heading an AdjP. In (ii), however, *angry* is part of a comparative in a DegP. These two possibilities are represented as (a) and (b), below.

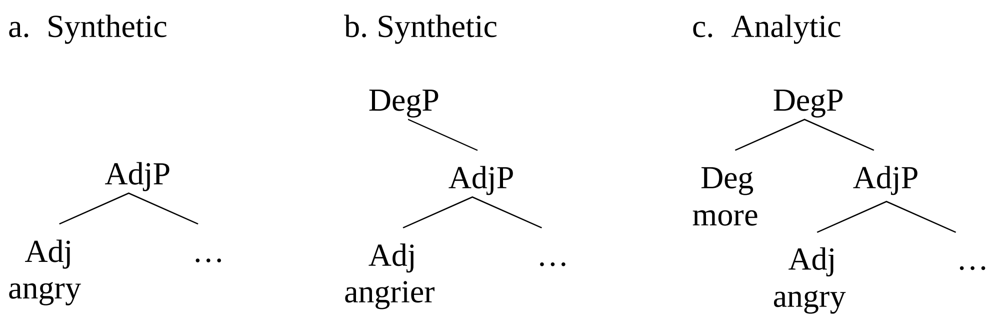


Figure 2.4: Comparative phrase-structure identification.

The problem for synthetic comparatives is that, at the point at which the stem is uttered, listeners may not have enough information to choose between the parses in (a) and (b). Choosing one parse before hearing the following disambiguating material potentially leads to garden paths, which could slow down processing. Compare this with the processing required to arrive at a successful parse of an analytic comparative, as in (c). The comparative marker *more* acts as an unambiguous signal that the following stem should be parsed within a DegP. Further, the positioning of the marker *before* the stem rather than after (cf. the synthetic *-er* suffix) allows for earlier recognition of the entire structure.

2.3.1 Design and Predictions

Both the WH and the PH make the same predictions with respect to listener preferences in complex environments. Under the WH, the analytic offers a processing advantage because it warns listeners of complexity, e.g. an upcoming infinitival complement. Under the PH, analytics are simpler to process because they avoid parsing difficulties associated with the synthetic. In order to differentiate between the WH and the PH, however, addressee behavior in simple environments must be considered.

Experiment 3 is a replication of Experiment 1, but with additional stimuli that allow us to measure addressee preferences in simple sentences, where no infinitival complement is present. I collect acceptability ratings of items that are manipulated according to the morphology type of the comparative (as in Experiments 1 and 2), as well as complexity. Example stimuli are shown below.

(3) Manipulating morphology type and complexity

a. Analytic-Complex: Kenny G is more mellow to listen to than Yanni.

b. Synthetic-Complex: Kenny G is mellower to listen to than Yanni.

c. Analytic-Simple: Kenny G is more mellow than Yanni.

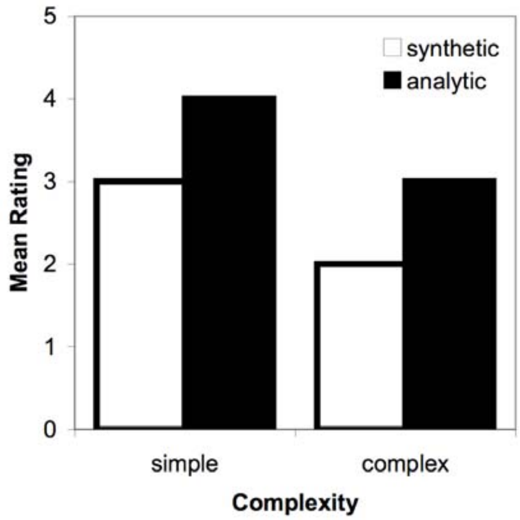
d. Synthetic-Simple: Kenny G is mellower than Yanni.

The experimental conditions represented by (3a) and (3b) have already been extensively studied in Experiments 1 and 2. Under conditions of complexity—i.e., when processing sentences containing a comparative followed by an infinitival complement—addressees tend to prefer the analytic according to both offline and online measures. Experiment 3 adds two new conditions in which addressees consider sentences that vary in comparative form, but do not contain an infinitival complement—(3c) and (3d). In sentences of this sort, which I refer to as *simple*, the comparative is always followed by a *than*-clause.

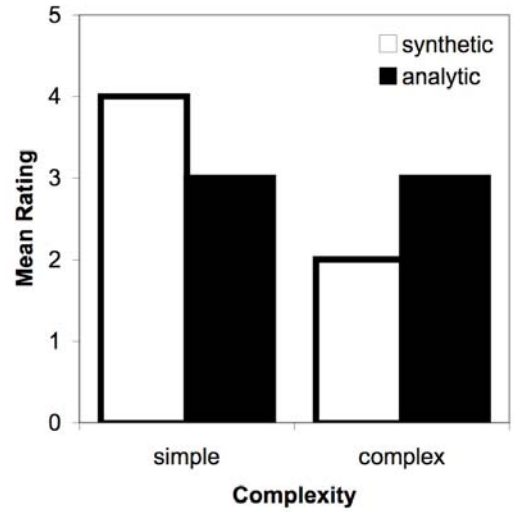
While the WH and PH converge in the predictions that they make for complex sentences, they make different predictions about the type of morphology that listeners will prefer in simple sentences. These predictions are sketched out in (4).

(4) Predictions of the Parsing and Warning Hypotheses

(a) Parsing



(b) Warning



According the PH, the analytic should be preferred in comprehension regardless of sentence complexity. We would therefore expect to see an analytic advantage in the simple condition that mirrors what has already been demonstrated in the complex condition, as in (4a). Under the WH, however, the advantage that the analytic enjoys in the complex condition is tied to its ability to predict complexity. The logic of the WH entails that analytic comparatives are associated with complexity, and that synthetic comparatives are associated with simpler sentences. Assuming that the WH is correct, these associations should lead to a flipping of preferences in the simple condition, where listeners' expectations are violated. Specifically, sentences in the analytic-simple condition should be dispreferred in comparison to sentences in the synthetic-simple condition because addressees should be expecting an infinitival complement, rather than a *than*-clause. This data pattern is shown in (4b).

In terms of the ANOVA, evidence in favor of the pattern in (4a)—the parsing hypothesis—would consist of a main effect of morphology type, and no interaction between morphology type and complexity. Evidence in favor of the warning hypothesis—(4b)—crucially depends on an interaction. A main effect of complexity is expected under both hypotheses.

2.3.2 Participants

100 UCSD undergraduates participated in Experiment 3. 51 of these individuals were students in an undergraduate course on language acquisition. The remaining 49 were recruited through the UCSD Psychology Department's subject pool. All participants received course credit for their time.

2.3.3 Materials

In order to increase the number of adjectives available for stimulus construction, the set of 18 that was used to create the stimulus sentences of Experiments 1 and 2 was supplemented with six new adjectives (*costly*, *deadly*, *frail*, *salty*, *unhappy*, *warm*). These six were chosen based on their ability in inflect either synthetically or analytically.

Critical sentences were constructed by embedding the 24 test adjectives into sentences that varied according to morphology type (analytic versus synthetic) and complexity (simple versus complex), as exemplified in (3). This process left us with 96 critical sentences, which were combined with 63 filler sentences to create four lists. Critical sentences are listed in Appendix B.

Each list contained 87 sentences total—24 critical sentences that were unique to the list they were placed in, and 63 fillers, which were identical across lists. Lists were constructed so that the four sentences associated with an adjective (see 3) were distributed one to each list. For the adjective *warm*, for example, the analytic-complex variant appeared in List 1, the synthetic-complex variant in List 2, the analytic-simple version in List 3, and the synthetic-simple version in List 4. Each list was balanced according to morphology type and complexity—i.e., lists contained exactly six exemplars from each of the four experimental conditions.

All lists began with five filler sentences. After this, critical sentences were pseudo-randomly inserted such that no two items from the same condition were adjacent, and all critical sentences were separated by two to three fillers. The overall ordering of fillers with respect to one other was determined randomly once; this order was then repeated over all four lists. This means that the first five sentences in each list were identical.

Lists were formatted and printed as pencil and paper questionnaires. The first page of each questionnaire consisted of a short survey on language background. This was followed by ratings instructions, and finally the questionnaire sentences. Each questionnaire sentence was followed by a 5-point, A-E ratings scale, with the *A* choice labeled *Excellent*, the *C* choice labeled *Average*, and the *E* choice labeled *Poor*. The questionnaire instructions directed participants to rate each sentence on the scale, and to record their answers on a scantron form.

2.3.4 Procedure

The 51 participants recruited through the language acquisition course were run in two large groups. The remaining 49 were run individually. All testing occurred in quiet rooms. Participants were randomly assigned to read a questionnaire based on one of the four lists described in §2.3.3.

Participants were handed a questionnaire and a scantron form, and were told to read through the questionnaire instructions thoroughly, and begin rating sentences. They were given as much time as they needed to complete the questionnaire (typically no longer than 30 minutes), and were supervised at all times by an experimenter.

After participants were tested, their responses were automatically recorded into a data file using a scantron machine. For purposes of data analysis, the A-E scores were converted to numeric values—5 for *A*, 4 for *B*, 3 for *C*, and so on.

2.3.5 Results

Permission to test students from the language acquisition course was contingent upon giving any individual who wanted to participate the opportunity to do so. This resulted in 17 non-native speakers of English being tested. Data from these individuals was excluded from subsequent analyses, leaving us with 83 native speakers of English. These were divided nearly equally among the four lists, with 22 participants reading List 1, 21 reading List 2, 19 reading List 3, and 21 reading List 4.

Altogether, native speakers of English read 1,992 critical sentences. Ratings for five of these were unavailable because the scantron machine either did not record a response, or recorded more than one response to the same sentence. Omitting these trials left 1,987 analyzable datapoints. An analysis of the distribution of ratings in this set shows that participants made full use of the entire scale. High ratings—either a 4 or 5—

were given to 40.4% of critical sentences; a medium rating of 3 was given to 21.8% of sentences, and a 1 or 2 was given to 37.8% of sentences.

When the data were considered according to the experimental manipulations, sentences in the analytic-simple condition fared the best (mean rating = 3.64), followed by a ratings tie between sentences in the synthetic-simple and analytic-complex conditions (both 3.11). Sentences in the synthetic-complex condition fared the worst (2.50). These averages were submitted to two ANOVAs, which treated morphology type (analytic versus synthetic), and complexity (simple versus complex) as repeated-measures variables, and modeled subjects (F_1) and items (F_2) as random variables. The results of these analyses show a main effect of morphology type, $F_1(1, 82) = 44.47$, $p < .0001$, $F_2(1, 23) = 13.10$, $p < .002$, a main effect of complexity, $F_1(1, 82) = 143.37$, $p < .0001$, $F_2(1, 23) = 23.67$, $p < .0001$, and no interaction, $F_1(1, 82) = 0.45$, $p = .50$, $F_2(1, 23) = 0.26$, $p = .61$. These findings are summarized graphically below.

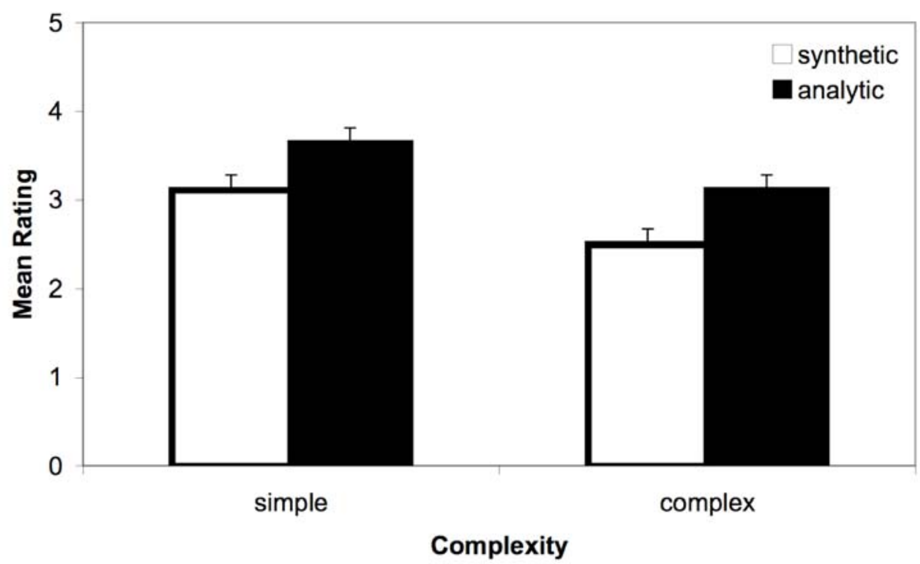


Figure 2.5: Effects of morphology type and complexity on acceptability

Note above that the analytic preference documented in Experiments 1 and 2 for sentences with infinitival complements (i.e., sentences in the complex condition) is replicated once again. The new finding contributed by Experiment 3 is that this same preference holds in simple sentences as well: the analytic is preferred in comprehension regardless of sentence complexity. These data thus come out strongly in favor of the PH. Under the WH, one would expect a significant analytic advantage in complex sentences as a result of the analytic's use as a conventionalized warning of complexity. This advantage would be attenuated or nonexistent in simple sentences because the lack of complex material following the comparative would go against listener's expectations, leading to processing difficulties and lower ratings. There is no hint of that kind of pattern in this data, however. Instead, the present findings are consistent with the notion that the analytic offers a processing advantage to listeners because it allows for earlier phrase-structure recognition, and avoids a temporary parsing ambiguity associated with the synthetic.

2.4 Discussion

I begin the discussion section by addressing a potential criticism. It might conceivably be argued that the significant effects of morphology type found across all three experiments in this chapter are the result of a frequency confound. According to this account, subjects show a preference for the analytic form not because it is simpler to process, as the parsing and analytic advantage hypotheses claim, but because the adjectives that were used to create stimulus sentences happened to inflect analytically more often. Speakers tend to prefer higher-frequency forms over lower-frequency forms, and the effect of morphology type in comprehension might simply reflect the fact that many of the test adjectives are more frequent in the analytic than the synthetic. This contention is rejected on multiple grounds, however. First, it is unlikely that adjectives that are more frequent in that analytic, i.e. that are *more*-dominant, drove the results of any of the experiments reported here. Of the 18 adjectives used in Experiments 1 and 2, only 5 have higher corpus counts in the analytic form than the synthetic. And of the 24 adjectives used in Experiment 3, only 8 are *more*-dominant. If anything, Experiments 1-3 appear to show that morphology type trumps frequency, since significant effects of morphology type were found using adjectives that were *not* favored in *more*.

Second, the results of all three experiments indicate that the preference for analytics was relatively *general*. In Experiments 1 and 2 respectively, 15 out of 18 and 13 out of 18 adjectives showed a preference for *more*. In Experiment 3, this proportion was 20 out of 24. The fact that some items that were *more*-dominant may have gone more strongly in the direction predicted by the analytic advantage hypothesis than non-*more*-dominant items does not indicate a frequency confound. It shows, rather, that frequency plays a role in analytic inflection. Experiments 1-3 were not designed to address frequency effects in comparative comprehension. I fully expect, however, that a morphology type manipulation combined with *more*-dominance might have a synergistic effect, and lead to very strong *more* preferences for some adjectives. Likewise, it is reasonable to assume that adjectives that are more frequent in the synthetic than the analytic will show a reduced, or nonexistent preference for the analytic. But none of this takes away from the fact that the participants in Experiments 1-3 tended to prefer the analytic in the majority of adjectives they were tested on, even when few of these were *more*-dominant. The present results thus suggest that the parsing and analytic advantage hypotheses are correct, and that—all other factors being equal—the analytic is preferred in comprehension.

I now turn to a more detailed interpretation of the results of Experiments 1-3. In her discussion of the parsing hypothesis, Mondorf (2003) cites Hawkins' theory of Early Immediate Constituents, or EIC (Hawkins, 2000, 2003). The EIC's central premise is that the human parser prefers linear orders that make more phrase structure recognizable earlier, rather than later. Hawkins (2000) demonstrates the validity of this principle in a corpus analysis of ordering preferences for manner, place and time adverbials in English. Verbal modifiers like these tend to be ordered shortest first and longest last. The result is that the overall phrase structure of the clause can be recognized in fewer words. This is demonstrated below.

(5) Early constituent recognition (Hawkins, 2000)

a. The astronomer [$_{VP}$ gazed [$_{PP1}$ through the telescope] [$_{PP2}$ into the dark but moonlit sky]].
1 2 3 4 5

b. The astronomer [$_{VP}$ gazed [$_{PP2}$ into the dark but moonlit sky] [$_{PP1}$ through the telescope]].
1 2 3 4 5 6 7 8

The sentences in (5) are identical except for the relative ordering of the two prepositional phrases, PP_1 and PP_2. These phrases differ in length: PP_1 is shorter than PP_2. When PP_1 occurs before PP_2, the structure of the VP is recognizable after just five words. When, however, PP_2 occurs before PP_1, full recognition of the VP structure is delayed until the eighth word. Hawkins shows that short-before-long orderings are preferred in naturally-occurring sentences, and further argues that this pattern is functionally motivated.

Minimizing recognition domains (i.e., making them smaller, as in 5a) facilitates processing by reducing the number of computations that are necessary for the online construction of syntactic and semantic representations (Hawkins, 2000: 257-258).

Hawkins' insights inform the present results in a quite straightforward manner. The data from all three of the experiments reported in this chapter are consistent with the parsing hypothesis, which predicts that analytic comparatives should be preferred over synthetics because they (among other things) allow for earlier recognition of the structure of the Degree Phrase. The comparative markers (*-er* and *more*) and the adjective stem can be variably ordered much like the PPs in (5). If the marker occurs before the stem, as in an analytic, then the structure of the DegP is available as *more* is being uttered. If, however, the marker occurs after the stem, the correct phrase structure may not be fully recoverable until *-er* is uttered. This difference probably amounts to an extremely short delay for synthetic comparatives. It may, however, be enough to drive the analytic preference that we see in the data.

The results of Experiments 1-3 are also informed by surprisal theory (Hale, 2001; Levy, in press). Given a partially parsed string, the processing difficulty associated with syntactically and semantically integrating the next unparsed word is directly proportional to its probability of occurrence in the context of the existing parse. Simply put, less probable words are harder to process. Such words have higher *surprisal*, which is a word's negative log probability. Levy (in press) demonstrates the utility of surprisal in an analysis of reading times from a study by Konieczny (2000), who looked at the processing difficulty associated with reading clause-final verbs in German embedded clauses. Relevant examples from Levy are presented below.

(6) Manipulating pre-verbal PP length

a.

Er	hat	den	Abgeordneten	begleitet,	und
He	has	the	delegate	escorted,	and

‘He escorted the delegate, and…’

b.

Er	hat	den	Abgeordneten	ans	Rednerpult	begleitet,	und
He	has	the	delegate	to_the	lectern	escorted,	and

‘He escorted the delegate to the lectern, and…’

c.

Er	hat	den	Abgeordneten	ans	große	Rednerpult	begleitet,	und
He	has	the	delegate	to_the	big	lectern	escorted,	and

‘He escorted the delegate to the large lectern, and…’

The sentences in (6) differ in terms of the length of the PP preceding the verb, *begleitet*. Because (6a) has no preceding PP, the PP length is conceived of as zero. In (6b) and (6c), however, the length of the PP is two words and three words, respectively. Konieczny (2000) found that reading times for verbs in clauses like those in (6) decreased as PP length increased. While this result is not predicted in other prominent theories of sentence comprehension (e.g., Dependency Locality Theory; Gibson, 2000), it is straightforwardly dealt with in a surprisal model.

Levy (in press) shows that this outcome is a function of how an upcoming word’s probability distribution shifts with knowledge of a sentence’s constituent history. The word order in German embedded clauses is verb-final. Because of this feature of German syntax, once listeners have recognized that they are processing an embedded clause, they know that a verb will be appearing soon. Since the clause cannot go on indefinitely, the probability that the next word will be a verb increases with each non-verb that is parsed. In (6a), for example, at the moment before the listener hears the verb, their current parse consists of a clause that already has a subject (*Er*) and an object (*den Abgeordneten*). These constituents are unlikely to appear twice within the same clause, and so are effectively out of the running as candidates for the upcoming constituent. Likewise, if a listener hears a sentence containing a goal PP (*ans Rednerpult*, or *ans große Rednerpult*), as in (6b) or (6c), this reduces the probability that another PP of the same nature will occur in the same clause. As each new pre-verbal word is parsed, the candidate set for the upcoming word shrinks, and the likelihood of a verb increases. Listeners are thus able to process the verb faster when it is preceded by longer PPs because they are more likely to expect that a verb will be next.

This same logic can be applied to the present results. The structure of English allows *more* to occur mainly as an adjectival (*more intelligent*) or nominal (*more cookies*) modifier. These two uses have very different distributions, which means that when comparative *more* is uttered, listeners are quite likely to interpret it correctly as an adjectival modifier. Once *more* is parsed in this manner, English listeners are in the same situation that German listeners are when they realize that they are processing an embedded clause—they know that a word from a certain class (i.e., an adjective) will be appearing soon. Compare this with the information that a listener has after having parsed the stem of a synthetic comparative. As noted previously, synthetic processing is associated with a temporary ambiguity that occurs before the *-er* suffix is processed. At that point in processing, the probability distribution for the upcoming constituent might favor *-er*. It might also, however, favor a large number of continuations that are consistent with a simple bare stem. A listener processing the same adjective in analytic form does not face as large an array of choices. The presence of *more* in a sentence's constituent history strongly predicts that the next word will be an adjective. Essentially, the structure of the analytic helps to avoid uncertainty in comparative processing, which leads to faster processing times, and an overall analytic preference.

2.5 Conclusions

The results of Experiments 1-3 are consistent with the notion that analytic comparatives are simpler for listeners to process, and that the analytic preference is due to a parsing advantage that the analytic form has over the synthetic. These findings are expected under at least two prominent theories of sentence comprehension—Hawkins' EIC (2000; 2003), and surprisal theory (Hale, 2001; Levy, in press).

Mondorf (2003) originally proposed that analytic comparatives are easier to process as a means of explaining production results. Recall that she found that speakers increased their rate of use of the analytic in complex environments. This was interpreted by Mondorf as evidence of an audience design strategy (V. S. Ferreira & Dell, 2000; Haywood et al., 2005; Kraljic & Brennan, 2005; Temperley, 2003), according to which speakers attempted to ease the processing burden on listeners by producing the form of the comparative that was preferred in comprehension. While the current data do indeed demonstrate that listeners prefer the analytic, they say nothing about whether speakers actually increase their use of the analytic in complex environments, or why they might do so. These question will be dealt with thoroughly in Chapter 3.

3. Production[7]

English provides two basic patterns for comparative inflection. A speaker can choose to mark an adjective for the comparative analytically using *more*, or synthetically using *-er*. But what determines speaker choice of comparative form? According to a number of prominent linguistic analyses (Embick, 2007; Poser, 1992; Stump, 2001), the production of one pattern over the other is determined almost wholly by reference to a single variable: the length of the adjective stem. Stem length (in syllables) and inflectional class are related such that longer stems are realized in the comparative analytically, and shorter stems are realized synthetically. This type of *univariate* treatment of production contrasts starkly with *multivariate* models. In line with the latter, recent statistical analyses over corpus data suggest that production choices ranging from the use of *that* as a sentential complementizer, to the decision to realize a ditransitive event as either a prepositional dative or a double object dative, are sensitive to a large number of factors (Bresnan et al., 2007; Jaeger, 2006; Roland et al., 2006).

The question of whether speaker knowledge is best analyzed as relying on one variable or many has enormous ramifications for linguistic theory. If univariate models are sufficient for describing speaker behavior in a domain such as comparative inflection, then linguistic competence is conceivably consistent with an underlying system of category structure in which category membership is discrete and deterministic, as has been traditionally assumed in linguistics (Aarts, 2004; Frege, 1960). That is, if all variability in formal expression is due to stem length, with short adjectives falling into

[7] The experimental work described in this chapter was planned and executed jointly with L. Robert Slevc.

the *-er* class, and long adjectives falling into the *more* class, then the comparative system should admit of a simple bipartite division, with each adjective stem assignable to only one inflectional class. But if multiple factors are required to explain speaker choice of comparative form, then competence would seem to rely on gradient, probabilistic category structure. Given that multiple weighted factors can be in conflict with one another, it is less likely that the dividing line between synthetic and analytic inflectional classes would be categorical, or that a single stem would always inflect in the same class.

There is already considerable evidence in favor of multivariate models of comparative inflection. Besides stem length (Aronoff, 1976; Embick, 2007; Frank, 1972; Poser, 1992; Stump, 2001), it has been demonstrated that speakers' choice of comparative form is sensitive to the identity of the stem's end segments (Aronoff, 1976; Frank, 1972; Graziano-King, 2003), frequency and gradability (Graziano-King, 2003; Graziano-King & Cairns, 2005), the stem's syntactic distribution (Díaz, 2006), and various sorts complexity (Mondorf, 2002, 2003). We address the contributions of this last factor in detail throughout the rest of this chapter.

Mondorf's (2003) basic finding is that, for adjectives that are able to participate in both inflectional classes, the rate of analytic use goes up when the processing environment is more complex. This is claimed to occur as a means of helping listeners. In Chapter 2, I tested whether listeners do indeed benefit from speaker use of the analytic, and found that sentences containing analytic comparatives were rated higher in acceptability (Experiment 1), and read faster (Experiment 2) than the same sentences with synthetics. I additionally presented data (Experiment 3) suggesting that the analytic comprehension preference is related to parsing: the analytic is simpler to parse because it avoids a temporary phrase structure ambiguity that afflicts the synthetic, and allows for earlier phrase structure recognition (Hawkins, 2000). In the current chapter, we shift our focus to production, and to a detailed critique of Mondorf's (2003) complexity finding. To the extent that speakers do take complexity into account when determining which comparative form to use, Mondorf's result strengthens the view that comparative inflection is sensitive to more than just the length of adjective stems, and further situates it among the increasingly large set of linguistic phenomena pointing to the probabilistic nature of grammatical knowledge. A secondary goal of the current chapter is to evaluate the functional explanation offered by Mondorf (2003) for increased use of the analytic in complex environments, namely that it represents an effort on the part of speakers to help listeners.

3.1 The Complexity Hypothesis

We refer to the proposition that complexity helps to determine the forms that speakers use as the *complexity hypothesis*. This idea has considerable support in the literature (Hawkins, 2000; Rohdenburg, 1996; Wasow, 1997), and was, in fact, addressed previously by Mondorf (2002) in terms of the effects that the presence of prepositional complements have on comparative form, e.g. *Adrian's father was more proud* <u>*of her*</u> *than ever before*. We are thus quite sympathetic to Mondorf's (2003) finding that speakers might tend to use the analytic at a higher rate in the presence of an infinitival complement. We believe, however, that the data on which this conclusion relies are not as compelling as they could be.

On the one hand, none of the 28 adjectives that Mondorf (2003) studied go against the hypothesis that there is increased use of the analytic when in the presence of an infinitival complement (i.e., in a complex environment). However, 11 of these are arguably irrelevant, since they are *never* followed by an infinitival complement in the corpus. And an additional 9 go in the predicted direction, but occur with an infinitival complement so few times that any conclusions that are drawn must be considered to be rather tentative. In the figure below, we review results for 11 of the 17 adjectives that are cited as evidence in favor of a correlation between complexity and comparative form. The white bars represent the proportion of times in the corpus that a comparative that was *not* followed by an infinitival complement was realized analytically—what we refer to as the *simple* condition. The black bars represent the proportion of times in the corpus that a comparative that *was* followed by an infinitival complement was realized analytically—the *complex* condition.

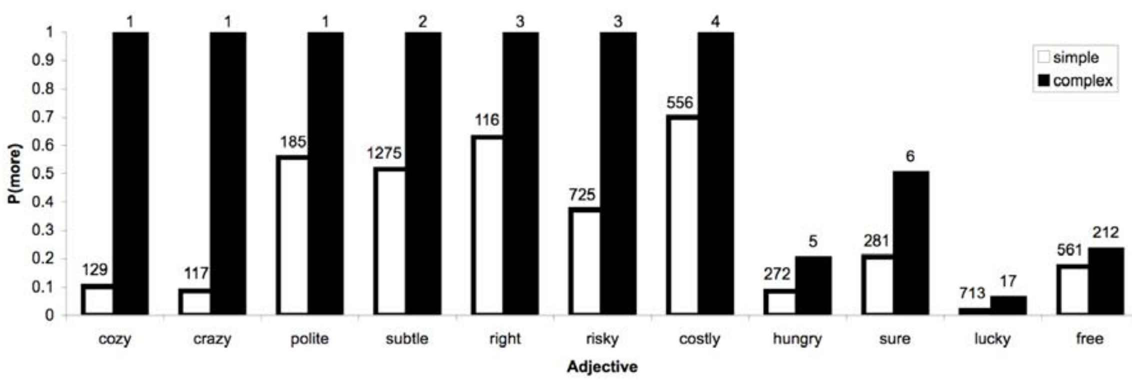


Figure 3.1: Low frequency counts in Mondorf (2003)[8]

[8] Many thanks to Britta Mondorf for generously providing us with the numbers needed to accurately reconstruct the data pattern illustrated in this figure.

Note that in every case above, the black bar is higher than its attendant white bar: every adjective has a higher proportion of analytic use in the complex condition. This would appear to go in favor of the complexity hypothesis.

The proportions in Figure 3.1, however, can be somewhat misleading. Some are based on very little data. Note the numbers printed above each bar. These represent the number of tokens from the corpus that were used to estimate the height of the bar. On the far left, for example, the 129 printed above *cozy*'s white bar indicates that there were 129 instances of comparative *cozy* in the corpus that were *not* followed by an infinitival complement. 13 of these 129 were inflected analytically, resulting in a roughly 10% rate of analytic inflection in the simple condition. The 90% difference between *cozy*'s simple and complex conditions appears to reflect a robust effect of syntactic complexity on comparative form. That is, however, until we note the number of tokens that went into estimating the rate of comparative inflection in the complex condition—i.e., the number of tokens that were used to estimate the height of *cozy*'s black bar. As it turns out, there is only one instance in the corpus of comparative *cozy* followed by an infinitival complement. Because this single token is analytically inflected, it gives us a 100% rate of analytic use in the complex condition.

If we treat the number printed above each bar as a measure of how confident we are in the estimated height of that bar, then we should have very little confidence in many of the estimated heights. This is particularly problematic for the first seven adjectives shown in Figure 3.1: *cozy*, *crazy*, *polite*, *subtle*, *right*, *risky* and *costly*. This set apparently demonstrates large effects of complexity. In all cases, however, we are relying on very little evidence to estimate the probability of analytic inflection in the complex condition. As we move from left to right in Figure 3.1, note that when larger numbers of tokens are available in the complex condition—for *hungry*, *sure*, *lucky* and *free*—the difference between the simple and complex conditions generally becomes much more modest. For the two adjectives with the largest numbers of tokens in the complex condition—*lucky* and *free*—the effect size is reduced to around 5%.

We make two comments concerning these data. First, the problem of low token frequencies is clearly acknowledged in Mondorf (2003). The only thing we have done in our presentation is make the ramifications of low frequency counts more explicit. Second, it is not our position that the low token frequencies given in Figure 3.1 invalidate Mondorf's (2003) conclusion concerning an increased use of the analytic in complex environments. Instead, we argue that the strength of such a conclusion is considerably weakened when the frequency numbers are taken into consideration, particularly with respect to the size of the complexity effect.

3.2 Semantic Complexity

While the main focus in Mondorf (2003) is on the effect that the presence of infinitival complements has on comparative form, Mondorf also devotes a significant proportion of time to cataloguing how comparative form is affected by various other types of complexity. For example, based on the assumption that abstract adjective senses are more complex than concrete adjective senses, Mondorf claims that semantic complexity correlates with increased use of the analytic. This is demonstrated in the invented sentences below.

(1) Semantic complexity

a. Simple (concrete): The Mediterranean is deeper than the Caribbean.

b. Complex (abstract): Keats is more deep than Shelley.

Sentence (1a) makes use of a concrete sense of the adjective *deep*, where *deep* refers to a physical unit of measure. The sentence in (1b) differs because it relies on an abstract sense of *deep*: the statement that Keats is deeper than Shelley refers to the relative degree of profundity associated with Keats' poetry as compared to Shelley's, rather than to the concrete notion that the man Keats might possess greater extension than Shelley along one of the three spatial dimensions. We call this kind of difference in complexity—i.e., the difference between concrete and abstract adjective senses—*semantic complexity*. The use of concrete adjective senses is semantically simple, and the use of abstract adjective senses is semantically complex. In the remainder of this chapter, we juxtapose semantic complexity with *syntactic complexity*, which refers to the presence or absence of an infinitival complement, as discussed in the preceding section, and in the comprehension studies of Chapter 2.

Mondorf's (2003) claim regarding sentences like those in (1) is that speakers tend to show increased use of the analytic in semantically complex sentences, like (1b). This possibility is supported using evidence from a single adjective in her 914 million-word corpus. *Remote* can be used in a concrete manner to describe physical distance. Its abstract usage, however, describes metaphorical distance, as when two political positions are remote with respect to one another on an ideological spectrum. According to the corpus data, *remote* does occur in the analytic more often when its sense is abstract (i.e., complex). This is illustrated in Figure 3.2.

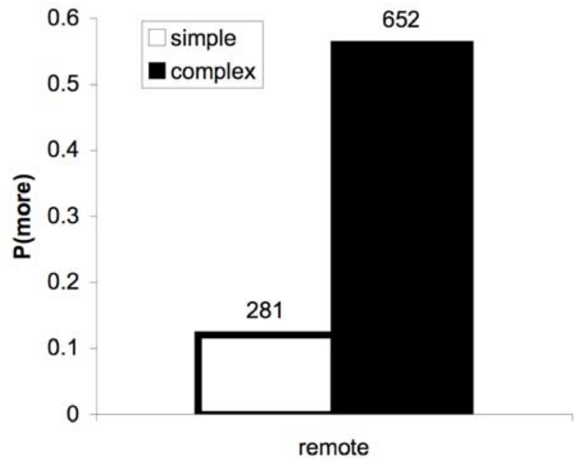


Figure 3.2: Effects of semantic complexity in Mondorf (2003)

We take the same position with respect to semantic complexity in Mondorf (2003) that we do for syntactic complexity: it is entirely possible that both kinds of manipulations lead to increased use of the analytic, and therefore serve to bolster the complexity hypothesis. In both cases, however, we feel that the corpus evidence is not as strong as one would like. The criticism that we level at the semantic complexity claim is different than the one we describe in §3.1 for syntactic complexity. It is not the case here that we are skeptical that there are enough tokens to come up with reliable estimates of the true probability of analytic inflection in each condition. The token numbers above each of the bars in Figure 3.2 appear to be reasonably high. Rather, the problem is that demonstrating an effect of semantic complexity in one adjective does not prove that semantic complexity is a general effect, which is the more interesting finding.

3.3 An Experimental Test of the Complexity Hypothesis

The problems identified in the preceding sections can be easily remedied in an experimental test of the complexity hypothesis. The general idea here is to attempt to replicate Mondorf's (2003) results by devising a method for eliciting different kinds of sentences containing comparatives. This method permits us to manipulate both syntactic and semantic complexity, and to track any resultant changes in speakers' preference for the analytic. We outline such a method in §3.5.

An experimental design in which comparatives are elicited with and without infinitival complements is able to circumvent the problem of low token frequencies described in §3.1 by allowing us to collect many tokens overall, and roughly equal numbers of tokens in the two conditions of interest. Likewise, experimentation allows for a simple means of ascertaining whether Mondorf's (2003) semantic complexity effect is limited to a single adjective—*remote*—or whether it applies more broadly. All that is needed is to test subjects using multiple items—something that is a standard practice in psycholinguistic methodologies. Data collected in this manner serves two purposes. First, it can either corroborate or contravene the results of Mondorf's (2003) corpus study. And second, in the case of corroboration, it can give us an estimate of effect size that is probably more reliable than what is available through the consideration of corpus data alone.

An additional argument in favor of experimental replication has to do with the assumption in Mondorf (2003) that the distribution of comparative forms is driven by processing-related factors. Her specific proposal is that the analytic is used as a compensatory strategy to ameliorate complexity-induced processing difficulty. While it is perfectly reasonable (and ultimately, probably correct) to assume that the distribution of analytic and synthetic variants is sensitive to processing concerns, corpus data are not ideally suited to investigating such claims. The problem lies in the complex web of correlations that exists in naturally-occurring data. In an experiment, relatively convincing arguments can often be made about the causal connection between independent variables and a dependent because there is a high degree of control. While commensurate control can be achieved in corpus studies that use advanced statistical techniques (Bresnan et al., 2007; Jaeger, 2006; Roland et al., 2006), this was not a part of Mondorf's (2003) method. The result is that there are potentially uncontrolled-for confounds that could be used to argue against the causal connection between use of the analytic and processing demands. In a controlled experimental design like the one we describe below, however, we are in a much stronger position to make processing claims because sentences with comparatives are produced relatively quickly. This ties processing much more convincingly to the choice between synthetic and analytic variants because the speeded timecourse of the production task greatly diminishes the possibility that subjects are responding based on non-processing-related factors (like style—see Curme, 1931 and Jespersen, 1949, for example).

In sum, an experimental test of the complexity hypothesis is motivated because it overcomes three difficulties that limit the strength of Mondorf's (2003) conclusions. It provides a solution to the problem of low token frequencies, it allows us to determine whether semantic complexity is a general effect, and it reduces the potential for uncontrolled confounds.

3.4 Audience Design

If it turns out that complexity has a systematic effect on comparative form, then we must additionally confront the question of *why* such an effect might exist. Mondorf

(2003) hypothesized that speakers increase their use of the analytic variant in complex environments in order to assist listeners: use of the analytic is a compensatory measure aimed at mitigating the effects of complexity and the increased processing loads that accompany it. Unfortunately, this hypothesis was outside the main scope of Mondorf's (2003) investigation, and so was not empirically evaluated. It is possible, however, for us to test it as a part of the present study.

The general proposal that speakers anticipate their listeners' needs and formulate utterances with those needs in mind is known as the *audience design hypothesis*. In one sense, it is obvious that speakers tailor both the content and the form of their output to accommodate addressee knowledge states and expectations: the subdiscipline of linguistic pragmatics is largely built upon the recognition of this reality (Grice, 1989; Levinson, 1983, 2000). Grice's (1989) cooperative principle and its attendant conversational maxims, in fact, anticipate and encapsulate the idea of audience design: speakers are constrained by conventions of rational interaction that find specific linguistic expression in different languages. Evidence in favor of audience design can be found in a number of areas. Bilingual speakers who share a common language with a monolingual interlocutor will, for example, switch to the interlocutor's language in order to facilitate communication. Similarly, speakers will tend to linguistically differentiate perceptually similar objects in order to avoid ambiguity—e.g., referring to a triangle as a *small triangle* rather than simply a *triangle* when in the presence of another, larger triangle (V. S. Ferreira, Slevc, & Rogers, 2005; Keysar et al., 1998; Lane, Groisman, & Ferreira, 2006).

The real question for those who investigate audience design is not whether its exists: some semblance of audience design is obviously operative in the cases mentioned above, as well as many others. Rather, the focus is on how deeply and in what ways the consideration of listeners' needs permeates linguistic behavior. Recent results from the ambiguity avoidance literature indicate a lack of consensus about whether audience design affects low-level grammatical and phonological encoding processes. Studies in this area target the production of potentially ambiguous sentences as a way to tease apart whether linguistic features are produced in order to help listeners, or as a means of facilitating speaker processing. Since speakers are by default privy to the correct interpretation of their own utterances, it is assumed that increased (or stronger) production of disambiguating features in sentences that are ambiguous should count as evidence in favor of audience design.

Ferreira and Dell (2000), for example, looked at complementizer production in sentences in which the subject of the embedded clause (underlined below) might, in the absence of a complementizer, be temporarily misanalyzed as the object of the matrix clause (e.g., *I knew <u>you</u> had booked a flight for tomorrow*), and found that *that* was not used more often when the potential for ambiguity existed than when it did not (as in sentences containing case-marked pronouns: *I knew <u>I</u> had booked a flight for tomorrow*). Likewise, Kraljic and Brennan (2005) reported that, when speakers directed addressees to move objects around a shared visual display using syntactically ambiguous commands such as *Put the dog in the basket on the star*, they did not produce disambiguating prosodic cues more strongly when the display supported both interpretations of the command (and thus was situationally ambiguous) verses when the display was situationally unambiguous. Since speakers in these tasks were apparently insensitive to addressee needs in precisely those situations in which addressees needed the most help (i.e., in the presence of ambiguity), these results suggest that audience design concerns do not penetrate linguistic behavior to the level of grammatical and phonological encoding.

Other studies, however, have come to the opposite conclusion. Temperley (2003) examined object relative clauses in the Wall Street Journal (WSJ) corpus and found that object relatives that contained temporary syntactic ambiguities were also more likely to be introduced by a disambiguating relative pronoun or complementizer. For example, writers tended to use *that* more often in sentences in which the relative clause subject (underlined) was a plural or mass noun without a determiner—e.g., *The car* (*that*) <u>*companies*</u> *like*…—than in sentences in which the relative clause subject was a pronoun—e.g., *The car* (*that*) <u>*I*</u> *like*…. Note that in the first case but not the second, it is possible for the relative clause subject to be misanalyzed as the main clause subject when *that* is not used. Temperley interprets these data as evidence that the use of relative pronouns and complementizers is at least partially driven by audience design considerations. Other evidence in favor of audience design in grammatical encoding comes from Haywood et al. (2005) who—contra Kraljic and Brennan (2005)—found that speakers *did* make efforts to disambiguate syntactically ambiguous commands like *Put the penguin in the cup on the star*. This was done through the addition of *that* or some other disambiguating word to the sentence (e.g., *Put the penguin* <u>*that's*</u> *in the cup on the star*). Haywood et al. (2005) explain the difference between their results and Kraljic and Brennan's (2005) as resulting from a slight methodological modification: interlocutors in their study switched speaker-addressee roles on every trial. This may have led to increased awareness of the potential for ambiguity, which consequently allowed speakers to target the use of ambiguity-avoidance features to exactly those circumstances where they might be most helpful to addressees.

The phenomenon that is the focus of the current study—increased use of the analytic in complex environments—seems to provide a potentially interesting test of the hypothesis that audience design is operative in grammatical encoding. First, note that none of the examples addressed above deal with a choice between competing morphological alternatives. Instead, they largely deal with the choice between syntactic frames for expressing a message, whether these frames will contain optional function words, and how the resulting sentences will be realized prosodically. To our knowledge, the current study is thus the first attempt to experimentally test audience design as it applies to morphology. Second, there is already some reason to believe that audience design is operative in the choice between comparative variants. Prior studies have noted that the linchpin for any successful demonstration of audience design is that speakers target the use of potentially helpful linguistic features to those contexts in which they might actually be most helpful to listeners (Brennan & Williams, 1995; Kraljic & Brennan, 2005). If the complexity hypothesis is correct, as Mondorf's (2003) data suggests, then this criterion would seem to be met in the present case. The data from Chapter 2 suggest that listeners prefer the analytic variant because it allows for earlier constituent recognition, and avoids a temporary parsing ambiguity that accompanies use of the synthetic (Hawkins, 2000; Mondorf, 2003). It thus seems that speakers may be increasing their rate of analytic use in exactly the right place. The question that remains is whether they are doing this to assist listeners, or as a means of facilitating their own production processes.

3.5 Experiment 4

Experiment 4 aims to test two hypotheses. According to the complexity hypothesis, speakers should use the analytic variant more often when the sentences that they produce are *complex*, versus when they are *simple*. We designed a production task in which complexity is modulated in two ways. In the *syntactic manipulation*, speakers were asked to produce comparatives in sentences that are either syntactically complex (i.e., contain an infinitival complement), or syntactically simple (i.e., do not contain an infinitival complement). In the semantic manipulation, speakers were asked to produce comparatives in sentences in which the adjective sense is semantically complex (i.e., abstract), or semantically simple (i.e., concrete). There are thus two factors in the design: *manipulation* (syntactic or semantic), and *complexity* (simple or complex). If both kinds of complexity are relevant to the choice of comparative form—as suggested in Mondorf (2003)—then we should see increased use of the analytic in both manipulations. This outcome can be tested for in an ANOVA as a main effect of complexity, with no complexity-by-manipulation interaction.

In order to test whether audience design affects choice of comparative form, we nest the above factors within a *communicative environment* manipulation. This kind of manipulation has been used in the past as a test of audience design (V. S. Ferreira & Dell, 2000; Kraljic & Brennan, 2005), and essentially entails that half of the speakers in an experiment produce their sentences alone, with no particular addressee in mind, while the other half produce them for a co-present addressee as part of a communicative task. We refer to these two environments as the *no listener* and *listener* conditions. If speakers are sensitive to addressee needs in the way that Mondorf (2003) predicts, then the effects of complexity should be magnified in the listener condition. Speakers in the listener condition should increase the rate of analytic production in response to complexity above and beyond the rate of increase that is seen in the no listener condition. This data pattern amounts to an ANOVA interaction of communicative environment and complexity.

3.5.1 Participants

84 undergraduates were recruited from the UCSD Department of Psychology subject pool. All subjects were native speakers of English who participated in the experiment in exchange for course credit.

3.5.2 Adjectives

We identified a set of 26 adjectives that exhibited optional comparative inflection. Ten of these were adjectives that were studied in Mondorf (2003). The remaining 16 were selected from among a group of 402 adjectives that varied in comparative form according to BNC (2001) frequency counts. For each of these 402 adjectives, we calculated the probability of synthetic inflection in the corpus—P(*er*). This was done to facilitate the selection of 16 items that represented a range of P(*er*) values—from those that were near 0.5 (i.e., were highly variable in comparative form), to those that were closer to one or zero (i.e., were generally favored in either -*er*, or *more*).

3.5.3 Prompts

Using these adjectives, we constructed a series of prompts that were designed to elicit various kinds of sentences containing comparatives. Each prompt consisted of three sentences, and reported results from an ostensible survey.

(2) An example prompt from Experiment 4

Film directors rated the heights of actors that they had worked with.

60% of directors said that Tom Cruise is short.

97% of directors said that Danny DeVito is short.

All prompts set up an implicit comparison between the entities mentioned in the second and third sentences, e.g. between Danny DeVito and Tom Cruise in (2). The comparison is established according to a standard formula. The first prompt sentence always provides a brief context for the comparison. The second sentence introduces the first entity (e.g., Tom Cruise), and locates it on the low end of the scale used for the comparison (e.g., the scale of shortness). The third sentence introduces the second entity (e.g., Danny DeVito), and locates it on the high end of the scale used for the comparison. In (4), for example, participants generally interpret Danny DeVito as being shorter than Tom Cruise because a relatively higher proportion of directors attach the label *short* to Danny DeVito (97%) than to Tom Cruise (60%). Based on the information given in each prompt, participants are asked to make up a sentence comparing the entities that are mentioned. This typically results in a comparative being produced. Crucially, note that participants are free to form the comparative however they want: i.e., nothing in (4) indicates whether *short* should be inflected synthetically as *shorter*, or analytically as *more short*.

3.5.4 Prompt Pairs

We used our 26 adjectives to construct 30 prompt pairs for use in the experiment. Half of these (i.e., 15 pairs) participated in the syntactic manipulation. The other half participated in the semantic manipulation. In the syntactic manipulation, the complex half of each pair was a prompt that was designed to elicit a syntactically complex sentence, while simple half of each pair was a prompt designed to elicit a syntactically simple sentence.

(3) The syntactic manipulation

a. Complex

The Village People were upset.

16% of music fans thought the sailor was angry to hear that the band was breaking up.

76% of music fans thought the cop was angry to hear that the band was breaking up.

b. Simple

The Village People were upset.

16% of music fans thought the sailor was angry.

76% of music fans thought the cop was angry.

For the syntactic manipulation, *complex* (see 3a) is defined as the presence of an infinitival complement (e.g., *to hear that the band was breaking up*) following the adjective (e.g., *angry*) in prompt sentences two and three. Because there is no infinitival complement following *angry* in sentences two and three of (3b), it is classified as *simple*.

In the semantic manipulation, one half of each pair was designed to elicit a semantically complex sentence, and the other half was designed to elicit a semantically simple sentence.

(4) Semantic manipulation

a. Complex

Computer programmers were surveyed about user-accessibility.

15% of programmers said that PCs are friendly.

80% of programmers said that Macs are friendly.

b. Simple

Dog owners were asked about unique breeds.

15% of respondents said that pit bulls are friendly.

80% of respondents said that golden retrievers are friendly.

Semantically complex prompts were defined as those that contained a use of the adjective (e.g., *friendly*) that was abstract in nature. The use of *friendly* in (4a), for instance, is abstract because it is being applied to entities that are only figuratively friendly: in the context of computers, *friendly* means easy-to-use, not amiable in terms of personality. Semantically simple prompts, on the other hand, used a concrete sense of the adjective. In (4b), *friendly* is concrete because it describes the personality traits of animate entities—dogs. The difference between abstract and concrete senses of the adjectives used in our semantic manipulation was independently verified in a pilot study, where participants rated adjectives uses that we label complex (see 4a) as being more abstract (i.e., less concrete) than adjective uses that we label simple (4b).

Each pair thus consisted of two prompts that differed in terms of complexity. Half of pairs involved manipulations of syntactic complexity and the other half involved manipulations of semantic complexity. All prompts used in Experiment 4 are listed in Appendix C.

3.5.5 Lists

Four lists of 36 trials each were constructed. On every trial, participants read and responded to a prompt. Thirty of these trials made use of prompts that were members of pairs—as in (3) and (4)—that varied methodically in terms of manipulation (syntactic versus semantic) and complexity (complex versus simple). These are referred to as *critical* trials. Another six trials in each list consisted of *filler* prompts that were drawn from a separate set of 12, all of which were constructed according to the conventions discussed in §3.5.3. All fillers were designed to elicit a synthetic comparative response. This was accomplished by using adjectives that were heavily favored in the synthetic according to BNC frequency counts, and by embedding these adjectives in prompts that were syntactically and semantically simple. Since both our syntactic and semantic manipulations should—according to Mondorf (2003)—increase the use of analytic comparatives, these fillers were included as a bulwark against the possibility that speakers would begin to produce analytics across the board, regardless of complexity. Such an outcome would negatively affect our ability to detect an effect of complexity by reducing variability in the dependent measure.

List 1 was populated with prompts in the following manner. For filler trials, six prompts were randomly selected from the set of 12 total filler prompts. These prompts were assigned to occur on trials 1, 7, 13, 19, 25 and 31. All other trials were critical trials. For critical trials, we took one member from each of the 30 prompt pairs such that the resulting set of prompts was evenly balanced in terms of manipulation (syntactic versus semantic), and complexity (complex versus simple). These 30 prompts were pseudo-randomly ordered in the remaining free trial slots with the restriction that no specific combination of manipulation and complexity was represented on more than two consecutive trials. List 2 was created by taking the six filler prompts that were not used in List 1, and ordering them to occur on trials 1, 7, 13, 19, 25 and 31. The critical prompts used in List 2 were simply the members of the 30 prompt pairs that were left over after List 1 had been populated. This amounted to a set of 30 prompts that—as in List 1—were evenly balanced in terms of manipulation type, and complexity. Critical prompts in List 2 were ordered such that List 2 was the complement of List 1. If, for example, a syntactically simple prompt (like 3b) occurred as Trial 8 on List 1, then its syntactically complex counterpart (e.g., 3a) would occur as Trial 8 on List 2. Likewise, if a semantically complex prompt (like 4a) occurred as Trial 22 on List 1, then its semantically simple counterpart (e.g., 4b) would occur as Trial 22 on List 2. List 3 was constructed in the same manner as List 1, with List 4 as List 3's complement.

3.5.6 Procedure

56 subjects participated as *speakers*. These subjects read prompts like those described above, and produced sentences comparing the entities mentioned in each prompt. Another 28 subjects participated as *listeners*. They were paired with 28 of the 56 speakers in a *listener condition*, which will be described in detail in §3.5.8.

All speakers were assigned to read from two lists. Half of the speakers read from Lists 1 and 2, and the other half from Lists 3 and 4. Of the speakers who read from Lists 1 and 2, half of them read List 1 first, while the other half read List 2 first. Similarly, of the speakers who read from Lists 3 and 4, half of them read List 3 first, and the other half read List 4 first.

3.5.7 The Production Task

All speakers completed an 84-trial production task in which they read prompts and made up sentences. The first 12 trials were *practice* trials. The remaining 72 were *critical* trials that were read off two of the 36-item lists (either Lists 1 and 2, or Lists 3 and 4). Speakers were seated in front of a computer monitor in a quiet testing room, and were equipped with a microphone headset attached to a voice key, as well as a standard button box within easy reaching distance. In the practice portion of each experimental session, speakers read instructions on the computer about how to create sentences based on prompts. They were told that they would be viewing the results of various surveys. Their job was to make up sentences that compared the two entities that were mentioned in the survey statements that they read. Speakers were encouraged to formulate their sentences before starting to speak, since the prompt would disappear from the monitor one second after voice onset. This was done to prevent speakers from using the prompt as a guide during articulation. After reading through the instructions, speakers were given 12 practice trials, with standardized feedback that showed the kinds of sentences they were expected to create. If, after completing the practice session, speakers did not feel comfortable with the task, or if the experimenter judged that they had not adequately mastered the task, they were given the opportunity to repeat practice. Otherwise, the test portion of the experiment immediately followed. Test trials differed from practice trials only in that no feedback was given. Speakers advanced from one test trial to the next by performing a button press using the button box.

Both the practice and test portions of the production task were controlled using Psyscope (Cohen et al., 1993). Responses in all test trials were recorded using an Olympus DS-4000 voice recorder. We additionally collected production latencies for each test trial by measuring the time it took speakers to begin speaking after the onscreen appearance of a prompt.

3.5.8 Communicative Environment

Besides controlling the complexity of the utterances that speakers produced through the use of the syntactic and semantic manipulations, we additionally attempted to control the communicative environment. 28 of our 56 speakers participated in a *no listener* condition, in which they produced the required sentences without any particular addressee in mind. In the *listener* condition, on the other hand, a different set of 28 speakers were paired with 28 co-present addressees. These speakers were told to direct their utterances to their assigned addressee, and were encouraged to formulate them with the addressee in mind.

The procedure used in the no listener condition has already been described in §3.5.7. This same procedure was used for subjects in the listener condition, but in a modified form. Subjects in the listener condition were scheduled for testing in pairs. For each pair, we randomly assigned one subject to the speaker role and one to the listener role using a coin toss. As in the no listener condition, speakers were seated in front of a computer monitor in a quiet testing room, and were equipped with a microphone headset attached to a voice key, as well as a standard button box within easy reaching distance. This time, however, listeners were seated across a table from speakers with a keyboard in front of them, in a position the prevented them from seeing the monitor. Speaker instructions differed slightly from those outlined above in that speakers were additionally told that it was important for them to make their sentences easy to understand, as their listeners would be performing a comprehension task. Listeners were handed a separate, printed set of instructions that told them that they would be rating the acceptability of the sentences that their speakers uttered. Examples of highly acceptable, moderately acceptable, and unacceptable sentences were given, and listeners were shown how to rate these on a one (odd-sounding) to five (perfectly acceptable) scale by pressing one, two, three, four or five on the keyboard. Because we were interested in maintaining the speaker's belief that the listener's task was communicative, and because we did not want to draw speakers' attention to the fact that the form of their utterances was being evaluated, listeners were additionally instructed not to say anything about the ratings task to speakers. Both interlocutors were encouraged to maintain eye contact while sentences were being spoken.

Individual trials proceeded along the same lines as described in §3.5.6 with the exception that, after the speaker produced a sentence, the listener would input a rating for it using the keyboard. Once a rating was entered, speakers were allowed perform a button press on the button box to advance to the next trial. As before, the experiment was controlled using Psyscope, and responses in all test trials were recorded using an

Olympus DS-4000 voice recorder. For each test trial, we collected a listener acceptability rating, and a production latency that was measured from the onscreen appearance of the prompt to the onset of speaking.

3.5.9 Coding

All critical trials were transcribed in full from the voice recordings made during each experimental session. These trials were coded into one of three categories. *Other* trials were those in which speakers failed to use a comparative, used a non-canonical form (e.g., a double-marked comparative, such as *more angrier*), vacillated between one form or another (e.g., *more angry…er*, *I mean angrier*), or used an adjective other than the one they had been prompted with. Syntactically complex trials were additionally labeled *other* if the speaker failed to produce an infinitival complement. All trials in which the synthetic variant was used were coded as *-er*. Trials in which the analytic variant was used were coded *more*. The utterances from all non-other trials were coded for length in words, excluding disfluencies.

3.5.10 Results

Subjects participated in 3,360 critical trials. Of these, 395 (11.7%) were coded *other*, 1454 (43.3%) were coded *more*, and 1511 (45.0%) were coded *-er*. A disproportionate number of *other* codes (66%) occurred on trials in which participants were supposed to create syntactically complex sentences. This reflects a tendency to sometimes omit the required infinitival complement. All trials coded as *other* were excluded from subsequent analyses.

We examine results that bear on the complexity hypothesis in §3.5.11 and §3.5.12. Results that bear on the audience design hypothesis are discussed in §3.5.13.

3.5.11 Effects of Complexity and Manipulation Type on Production Latency

We have thus far applied the terms *simple* and *complex* to our stimuli without having an objective measure of whether the manipulations that we implemented actually modulate complexity. Intuitively, having to process an additional phrase—the infinitival complement—or having to deal with adjective senses that are more abstract represents an increase in complexity over what is required to process simple stimuli. In an attempt to supplement this intuition with a more conventionally agreed-upon measure of complexity, we conducted an analysis of production latencies from Experiment 4. Recall that we measured the time it took speakers to begin speaking by timing from the onset of the stimulus prompt to the onset of their utterance. Under the standard assumption that stimuli that take longer to process pose more cognitive difficulty, this now puts us in a position to predict that stimuli in Experiment 4 that occur in the complex conditions should show significantly longer production latencies that stimuli from the simple conditions.

The production latency analysis was conducted by first eliminating—on a subject-by-subject basis—latencies that were greater than two standard deviations from the mean. In order to control for utterance length, we then performed separate linear regressions of production latency on utterance length for each subject (F. Ferreira & Clifton, 1986). This has the effect of factoring out contributions from length to the production latency. The residuals from these analyses were subsequently entered into a 2 (complexity) × 2 (manipulation) × 2 (communicative environment) ANOVA, with complexity and manipulation as within-subjects factors, and communicative environment as a between-subject factor. Results show main effects of complexity, $F(1, 53) = 32.58$, $p < .001$, manipulation, $F(1, 53) = 17.76$, $p < .001$, and a significant interaction of complexity and manipulation, $F(1, 53) = 15.58$, $p < .001$. The complexity and manipulation findings are summarized below.

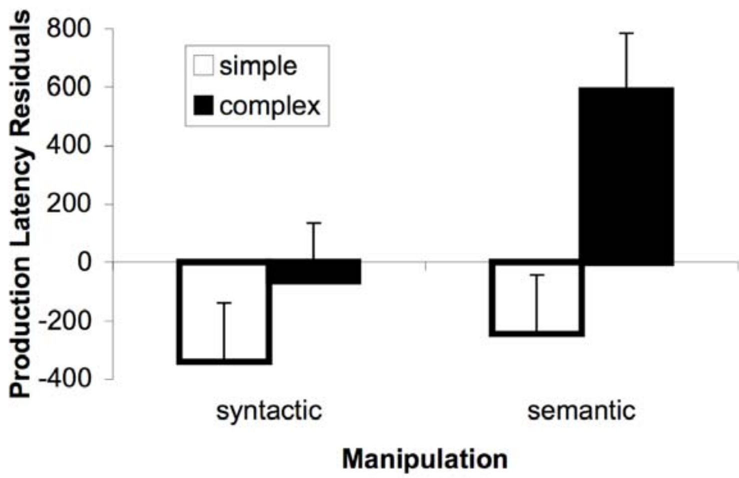


Figure 3.3: Effects of complexity and manipulation on production latency residuals

The figure above shows the relationship between mean production latency residuals—which can be thought of as deviations from the simple length-based predictions of the regression model—as a function of manipulation and complexity. Means that are close to zero indicate that utterance length is able to predict a good deal of the variation in production latency. Means that are further away, however, indicate that there is variation that cannot be accounted for solely by utterance length. For example, the black bar on the far right of the figure indicates that the mean production latency in the complex-semantic condition is 587 milliseconds *longer* than one would expect based on utterance length alone. Likewise, the white bar on the far left of the figure shows that the mean production latency in the simple-syntactic condition is 339 milliseconds *shorter* than would be expected based on utterance length.

Overall, the figure illustrates both the main effect of complexity—with the mean of the white bars (-291 ms) being reliably smaller than the mean of the black bars (262 ms)—and the main effect of manipulation—with the mean of the two syntactic bars (-201 ms) being reliably smaller than the mean of the two semantic bars (171 ms). Because different sets of items were used in each manipulation, manipulation type is confounded with item set. As a result, we refrain from reading too much into the main effect of manipulation. The significant interaction of complexity and manipulation indicates that subjects did not respond to the syntactic and semantic manipulations in the same manner. In order to rule out the possibility that one manipulation was driving the overall complexity effect on its own, we conducted pairwise *t*-tests between the simple and complex conditions within each manipulation type. These demonstrate that both manipulations led to statistically reliable differences: $t(53) = -2.80$, $p < .001$ for the comparison between the syntactically simple and syntactically complex conditions, and $t(53) = -8.43$, $p < .001$ for the comparison between the semantically simple and semantically complex conditions. This suggests that the complexity-by-manipulation interaction may be driven by a significantly larger effect size on the semantic side, although this conclusion must be tempered considerably by the fact that different items were used in the two manipulations.

We were surprised to find that the mean residual production latency in the syntactic-complex condition was so near zero. Values near zero tend to indicate little deviation from the predictions made by the purely length-based regression model. In order to ascertain whether the syntactic-complex mean was reliably different than zero, we constructed a 95% confidence interval around it. Results show that the interval contains zero (-63.47 ± 199). This suggests that syntactic complexity may be roughly isomorphic with utterance length. That is, there appears to be very little in our syntactically complex items that cannot be explained by appealing solely to length.

Because the manipulations that we apply in Experiment 4 reliably affect production latency, and because response latencies are a standard measure of task difficulty, we conclude that these manipulations really do modulate complexity. These results thus provide an independent measure of complexity that matches up satisfactorily with our intuitions.

3.5.12 Effects of Complexity and Manipulation Type on P(*more*)

We now turn to the effects of complexity and manipulation type on the probability that a comparative will be inflected analytically. Our results show a main effect of complexity, with speakers producing the analytic more often in the complex condition (0.53) than in the simple condition (0.46), $F(1, 54) = 31.32$, $p < .0001$, and a main effect of manipulation, with increased use of the analytic in the syntactic manipulation (0.51) over the semantic manipulation (0.48), $F(1, 54) = 5.68$, $p < .03$. Complexity and manipulation did not interact, $F(1, 54) = 1.76$, $p = 0.19$. These results are summarized graphically below.

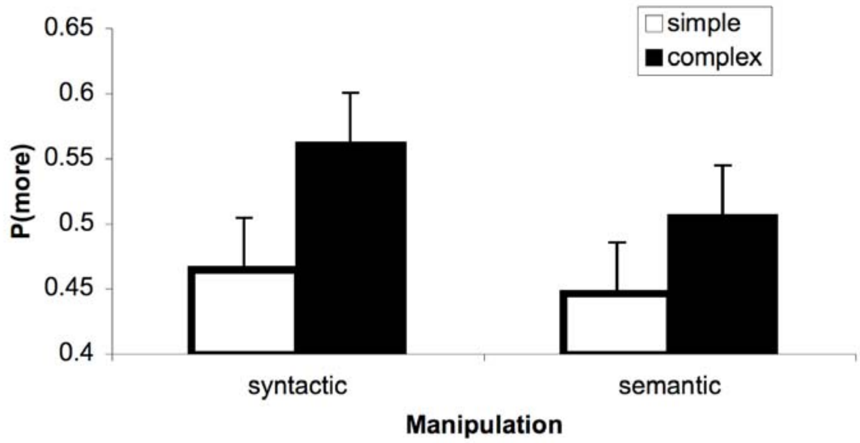


Figure 3.4: Probability of analytic inflection by complexity and manipulation type

The findings in Figure 3.4 replicate and reinforce the conclusions reached in Mondorf (2003). Both syntactic and semantic complexity can affect speakers' selection of comparative form, with increased complexity positively correlating with increased used of the analytic. Further, the lack of an interaction suggests that speakers treated both kinds of manipulations equivalently. That is, the main effect of complexity can be thought of as arising from the roughly equal contributions of the syntactic and semantic manipulations. As above, we refrain from interpreting the main effect of manipulation as a reflection of the greater difficulty of one type of manipulation over the other, since it is potentially the result of our use of different sets of items in the syntactic and semantic manipulations.

Overall, the findings from our analyses of production latency and probability of analytic inflection are mutually corroborative, and support Mondorf (2003). The corpus and experimental evidence agree that use of the analytic variant goes up in the face of complexity, and the results from §3.5.11 now allow us to define complexity using a standard metric of processing difficulty.

3.5.13 Effect of Communicative Environment

In order to investigate the effect of communicative environment, half of our 56 speakers produced their sentences with an interlocutor present (the *listener* condition) while the other half did not (the *no listener* condition). We included this factor as a between-subjects variable in both of the ANOVAs described in the preceding two sections. The results of these analyses show no main effect of communicative environment on either the production latency residuals, $F(1, 53) = 0.97$, $p < 0.40$, or the probability of analytic inflection, $F(1, 54) = 0.02$, $p = 0.90$. Additionally, there were no significant interactions of communicative environment with any other factors in either analysis. The figures below illustrate the lack of effect that communicative environment had on complexity: speakers treated complex items in the same way, regardless of whether a listener was present or not.

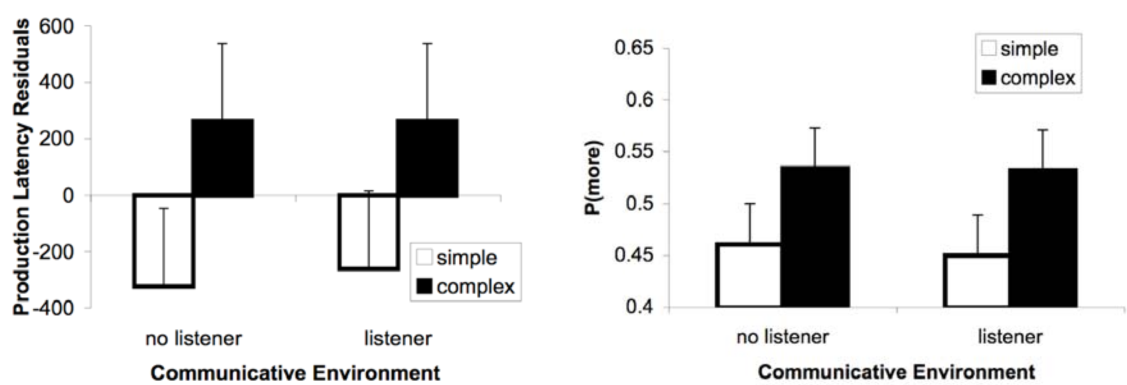


Figure 3.5: Effects of communicative environment on production latency and P(*more*)

If the choice between synthetic and analytic comparative variants were sensitive to listener needs, we might expect speaker behavior to vary in some way according to whether or not a listener was present. There is, however, no support for this hypothesis in the current data.

3.6 Discussion

We tested two hypotheses from Mondorf (2003) concerning the use of analytic and synthetic comparative variants in production. The first is the *complexity hypothesis*, which states that speakers should show an increased preference for the analytic variant in complex processing environments. We created two such environments—by manipulating the presence of infinitival complements and by modulating the abstractness of adjective senses in the sentences that speakers were prompted to create. The results indicate that speakers are sensitive to these manipulations, that our complex sentences really are more difficult to process, and that speakers responded to this increased difficulty by reliably increasing their use of the analytic. The second hypothesis that was tested concerned *audience design*. Is there evidence that speakers' choice between analytic and synthetic variants might be motivated by a consideration of listeners' needs? This was tested using a manipulation of the communicative environment: some speakers produced their sentences in front of a computer monitor, with no interlocutor present, while others produced them as part of a communicative task that involved a co-present addressee. This manipulation did not affect speaker production in any measurable way. Our results thus do not support the audience design hypothesis.

The position articulated in Mondorf (2003) that complexity affects comparative form is considerably strengthened based on the present experimental results. We provide within-subject experimental manipulations that avoid two basic problems that limited Mondorf's conclusions. First, with respect to syntactic complexity, we circumvented the problem of low corpus frequency counts for sentences containing comparatives followed by infinitival complements by devising a method for eliciting sentences with these features. This allowed us to collect many examples of sentences that would otherwise be quite rare in naturalistic use, and additionally enabled us to collect roughly equal numbers of sentences in the simple and complex conditions. Second, the effect of semantic complexity is placed on much firmer ground because we present evidence for it from a range of different adjectives, thereby establishing it as something that is likely to apply to all adjectives that have senses that vary in concreteness.

These complexity results are in stark opposition to a number of prominent linguistic analyses of the comparative (Embick, 2007; Poser, 1992; Stump, 2001), which treat the choice between the analytic and synthetic patterns of inflection as something that is driven almost entirely by syllabicity. Instead, our findings are in concord with a growing body of literature that suggests that comparative production is sensitive to a number of factors beyond mere syllabicity (Díaz, 2006; Graziano-King, 2003; Mondorf, 2002, 2003). Graziano-King (2003), for example, found that participants in ratings and forced-choice judgment tasks tended to disprefer inflecting certain monosyllabic adjective stems with -*er* (e.g., **apter*, **righter*)—an outcome that goes against the predictions of a syllable-only account of comparative inflection. Instead, suffixation with -*er* was found to be available only for high frequency, gradable monosyllabic adjective stems. Likewise, evidence on the Old English origin of the analytic pattern of inflection also suggests that a syllable-only account is insufficient. Díaz (2006) provides evidence that *more*'s early function in English was as an adverb, and that it was able to eventually become an adjectival degree modifier by transitioning from verbs to adjectives through participial forms like *bent* and *known*, which share properties in common with both verbs and adjectives. Note however that, in modern English, synthetic inflection is often unavailable for these forms (**benter*, **knowner*), even though they are monosyllabic. This indicates that the way in which speakers inflect a stem for the comparative is sensitive to the stem's syntactic distribution. Stems with verb-like distributions—i.e., participles—tend to be resistant to suffixation.

The fact that comparative production appears to be sensitive to multiple, sometimes conflicting constraints aligns it with recent probabilistic models of production (Bresnan, 2006; Bresnan et al., 2007; Jaeger, 2006; Roland et al., 2006). Such models view the knowledge that speakers deploy when using language as rich and multifaceted. This position contrasts starkly with univariate analyses of comparative inflection that have been proposed by theorists working in a wide array of grammatical frameworks (Embick, 2007; Poser, 1992; Stump, 2001). According to these analyses, comparative form is solely (or predominantly) determined by reference to a single factor—stem length in syllables. Because each adjective stem has a fixed length, this approach predicts that all adjectives will be inflected categorically as either *-er* comparatives, or a *more* comparatives, and that no optionality will exist. I provide a thorough critique of these approaches and their assumptions in Chapter 5. Here, it is sufficient to note that the highly dubious claim that optionality does not exist is convincingly refuted by the present data, and that a model of comparative production that makes reference to multiple, weighted constraints is more empirically adequate than one that is based on syllabicity alone.

Even though our test of the audience design hypothesis ended with a null result, and null results can be notoriously difficult to interpret, we are not particularly surprised that the communicative environment manipulation did not find evidence in favor of audience design. While it is true that our speakers increased their use of the analytic in exactly those circumstances in which it might be most helpful to listeners, this bit of evidence alone does not constitute unequivocal support for the audience design hypothesis. Keysar et al. (1998) point out that utterances that seem to be designed specifically for an addressee are like a well-fitting suit. It could be that the suit fits well because it was hand-tailored for the owner. Alternatively however, the suit might fit well for some other reason—e.g., it may have been designed for another individual who just happens to have the same measurements as the current owner. The significance of this analogy for the current case is that we do not know whether the behavior in question—increased use of the analytic in complex environments—is part of a strategy to "fit" utterances to listeners' needs, or if this same pattern might be the result of constraints on speakers' own production processes. This second possibility is particularly plausible given evidence from Ferreira and colleagues (V. S. Ferreira & Dell, 2000; V. S. Ferreira & Firato, 2002) that suggests that speakers might use optional *that* complementizers to buy time for the production of embedded clause material that is harder to access. Use of the analytic comparative variant could serve a similar function. Faced with having to a produce a complex sentence, speakers might conceivably use the analytic as a means of slowing down production in order to complete access and encoding processes that are more time-intensive as a result of the increased load. If true, this kind of story would account for the distribution of the analytic without the need for invoking audience design.

The lack of an effect of communicative environment in Experiment 4 is additionally unsurprising on architectural grounds. That audience design strategies would be robustly visible at the level of grammatical encoding is unlikely given that speakers are generally not good at the online identification of utterance characteristics that might pose difficulty for listeners. Ferreira et al. (2005), for example, show that speakers are often unaware of frank linguistic ambiguities—that *bat* might be used to alternatively describe a flying mammal or a stick used to hit a baseball, for instance. This observation is consistent with what Levinson (2000) calls the argument from design. According to this position, because production is inherently more time consuming than comprehension, linguistic communication has evolved in the direction of good listener inferencing abilities, and away from the kind of complexification of production that would occur if audience design were an integral component of grammatical encoding. Levinson argues that physiological constraints on the speed of articulation mean that prearticulation processes in production (Wheeldon & Levelt, 1995), as well as parsing and comprehension processes generally (Mehler et al., 1993), run three to four times faster than articulation. This constitutes a bottleneck in communication that has been overcome by shifting the processing burden onto listener inferencing. If addressees can construct speaker meaning by using pragmatic heuristics and other cues to constrain the possible interpretations of an utterance, then speakers can get around the articulatory bottleneck simply by saying less. Good listener inferencing also means that it is not absolutely necessary that speakers be acutely sensitive to listener needs. Communication will in most cases proceed successfully even when speaker utterances are linguistically ambiguous. Speakers only have to be clear enough for listeners to draw the right inferences.

This is not to say that speakers are completely incapable of incorporating audience design into low-level production processes. As noted above, findings from

Temperley (2003) and Haywood et al. (2005) appear to indicate that this is possible. We would argue, however, that because addressees are particularly adept at inferring the correct meaning of utterances (Levinson, 2000), audience design considerations might only be detectable at the level of grammatical encoding under special circumstances. Note that Temperley's (2003) results come from a corpus of heavily edited, written data—the WSJ corpus. It could be that the considerably slower rate of production and the fact that sentences are more often revised in the written modality acted to facilitate writers' ability to detect and repair potential ambiguities. Similarly, the positive audience design result in Haywood et al. (2005) might reflect their particular experimental design rather than something that robustly occurs in natural conversation. Participants in their task alternated speaker and addressee roles on every trial. This method was used specifically in an attempt to highlight the potential for ambiguity that existed in the sentences being produced. We argue that demonstrations of this sort only show that audience design strategies *can* be incorporated at the stage of grammatical encoding, not that audience design *routinely* plays a role in low-level production processes.

3.7 Conclusions

We find that speakers' choice of comparative form is systematically affected by complexity. Speakers appear to refer to multiple, weighted variables (not simply stem length) in deciding the form that their utterances should take. And while addressee needs may be considered in certain, circumscribed situations (Haywood et al., 2005; Temperley, 2003), there is no evidence that speakers in Experiment 4 chose between comparative alternatives based on anything other than the demands of their own production processes.

These findings, coupled with those discussed in Chapter 2, suggest that speakers and listeners have both converged on a preference for the analytic, but for very different reasons: listeners because it simplifies parsing, and speakers because it allows them to slow down production, thereby providing additional milliseconds for time-intensive access and encoding processes. The fact that the analytic is so roundly advantageous begs the question of why adjectives might inflect in the synthetic at all. This is addressed by Levinson (2000), who notes that the articulatory bottleneck should favor shorter forms over longer forms. This may be particularly true for adjectives that inflect synthetically most of the time. These forms are also likely to be high in frequency (Graziano-King, 2003), which should also facilitate production.

4. Acquisition

Structural priming studies have capitalized on the existence of optionality to investigate the nature of children's syntactic knowledge. For example, if children's early syntactic representations are conservative and item-based, as has been suggested by a number of researchers (Cameron-Faulkner, Lieven, & Tomasello, 2003; Tomasello, 2000, 2003), then children should fail to show priming effects. If the representations are sufficiently abstract, however—as is the case later in development—then priming effects should be observed. This finding has been borne out in a number of studies (Huttenlocher et al., 2004; Savage et al., 2003, 2006).

The abstractness of grammatical knowledge can be tested in other ways as well though. Traditionally, overgeneralization errors have been treated as evidence that children are operating on the basis of abstract linguistic representations. Overapplications of the English *-ed* past tense marker and the *-s* plural, for example, are standardly assumed to reflect the development of abstract generalizations that allow *-ed* to be suffixed to any verb, and *-s* to any noun. The same kinds of overuse errors are also studied in the acquisition of English comparative inflection.

In the current chapter, I look at overgeneralization errors in comparative acquisition as a means of determining if and when children develop abstract ADJ*er* and *more*+ADJ inflectional patterns. I begin by offering a critique of previous work in this area, then proceed to develop an experimental methodology that, I believe, is able to address certain areas of deficiency.

4.1 Overgeneralization in Comparative Inflection

Prior studies of comparative production agree that children go through a period of -*er* overuse (Clahsen & Temple, 2003; Gathercole, 1985; Graziano-King & Cairns, 2005). Gathercole, for example, worked with 56 children ages 2;6 to 5;9 and found that, averaged across all ages, -*er* was applied to adjective stems that adults generally inflect with *more* 28.6% of the time. Clahsen and Temple (2003) noted a similar trend: across twenty normally-developing children ages 5;1 to 7;ll, the mean rate of -*er* overuse was 41.6%. This number is nearly identical to the 44% rate of -*er* overuse reported by Graziano-King and Cairns (2005) for 29 children between 4;9 and 10;9.

According to at least two of these studies (Clahsen & Temple, 2003; Graziano-King & Cairns, 2005), the misapplication of -*er* to adjectives that adults normally inflect with *more* is significant because it counts as evidence of -*er overgeneralization*. That is, the systematic use of -*er* to form the comparative of adjectives such as *dangerous* (**dangerouser*) shows that children possess some kind of abstract ADJ*er* representation that allows for the suffixation of -*er* to *any* adjective stem. This is noteworthy in that it demonstrates that children's linguistic behavior cannot be accounted for by the rote memorization of stem-comparative pairs—a hypothesis that predicts no overuse errors. Instead, the misuse of -*er* shows that children are moving beyond rote memorization to the construction of an abstract system for doing comparative inflection—that is, a grammar.

Given the significance that is attached to -*er* overuse in these studies, it is somewhat surprising that no attention is paid to *more* overuse. If the misapplication of

-*er* counts as evidence in favor of the presence of an abstract ADJ*er* pattern of inflection, then *more* overuse (e.g., **more bright*, cf. adults' *brighter*) should count as evidence in favor of a *more*+ADJ pattern. In fact, if one looks closely at previous results, *more* overuse appears to occur at rates running from 3.6% (Gathercole, 1985) to 5.9% (Clahsen & Temple, 2003) to 16.4% (Graziano-King & Cairns, 2005). None of these authors, however, suggest that that *more* overuse errors might be indicative of a stage of *more* overgeneralization. Furthermore, no statistical tests are done to determine if children's rates of *more* usage differ from adults'. Note that 3.6%, 5.9% and 16.4% are all smaller than the documented rates of -*er* overuse: 28.6%, 41.6% and 44%. It could be that the relatively infrequent occurrence of *more* overuse has led researchers to dismiss it as noise, and consequently to disregard it as evidence of *more* overgeneralization.

It is the position of the current study, however, that dismissal along these lines would be in error. *More* overuse is to be expected during the course of comparative acquisition because children must, at some point in development, establish an abstract *more*+ADJ pattern of inflection. This reasoning is based on the notion that *more* is the default means of creating a comparative in the adult grammar. The *more* pattern is extended to a much larger number of adjectives than -*er* (Boyd, 2006), it attaches to all kinds of stems (unlike -*er*, which is phonologically restricted for the most part to monosyllabic stems and disyllables ending in -*y*), and it is not limited to high frequency stems like -*er* is (Graziano-King, 2003). Based on this set of criteria, the *more* pattern of inflection is the more general, default pattern in English (Pinker & Prince, 1994). Thus, in order to achieve the adult target, children must at some point develop an abstract representation that can be used to productively combine *more* with a wide variety of adjectives. If such a representation exits, one might expect it to manifest itself through overapplication, in much the same way that the ADJ*er* pattern is manifested through -*er* overuse.

4.2 Type Frequency and Generalization

A number of researchers have identified type frequency as a factor that contributes to the formation of linguistic generalizations (Goldberg, 2006; Plunkett & Marchman, 1991). The basic idea with respect to the learning of inflectional morphology is that, if the number of individual lexemes that children inflect in the same way is large enough, a higher-order generalization will be formed. In terms of comparative learning this means that children will form abstract ADJ*er*, or *more*+ADJ generalizations if they gain experience with enough -*er*, or *more* comparatives.

This simple model of generalization interacts with certain properties of the comparative domain to make testable predictions about what kinds of generalizations will be formed, and in what order. All else being equal, children will learn high frequency words before low frequency words. Since high frequency words tend to be shorter, this means that the initial set of words that children learn will also likely be short. These facts are of enormous consequence for comparative acquisition because short, high frequency adjectives tend to take -*er* (Aronoff, 1976; Frank, 1972; Graziano-King, 2003). Children's early vocabularies will thus contain relatively large numbers of -*er* adjectives, but fewer *more* adjectives. This asymmetry in class size—with children having significant early experience with -*er* types, but little experience with *more* types—predicts the early formation of an abstract ADJ*er* pattern of inflection. As their vocabularies grow, however, and come to include larger numbers of lower-frequency *more* adjectives, children will gain enough experience with the *more* pattern to induce the formation of an overarching *more*+ADJ generalization. Assuming that the existence of abstract generalizations can be detected through overuse errors, I am thus able to predict an initial period of -*er* overuse, followed later by a period of *more* overuse.

While these predictions are certainly consistent with the patterns of overuse documented in previous studies of comparative production (Gathercole, 1985; Clahsen & Temple, 2003; Graziano-King & Cairns, 2005), it is unclear whether the rates of *more* overuse that were reported are large enough (i.e., different enough from adult behavior) to reliably count as overuse. Put another way, no previous study has focused on *more* overuse, or performed the statistical tests necessary to determine if children really do overuse *more*. It therefore remains an open question whether comparative acquisition is characterized by one, or two patterns of overgeneralization.

4.3 Elicited Production and Response Perseveration

Prior research on comparative acquisition has relied on an elicitation method that is more or less identical across studies (Gathercole, 1985; Clahsen & Temple, 2003; Graziano-King & Cairns, 2005). Children participate in a series of *production trials* in which they see two items that differ on some noticeable attribute, and are prompted to finish a sentence describing the items with a comparative form. A child might, for example, be shown two trucks—one longer than the other—and be prompted with *This truck is long, but this truck is even* ______.

One problem with the way that this method has been deployed in the past, is that production trials typically follow one another back-to-back, without interruption. This is troublesome because it leads to a potential confound by increasing the possibility of response perseveration across trials. Consider, for instance, the sequence of trials given in Table 4.1.

Table 4.1: Response perseveration in elicited production

Trial	Production
$t - 1$	*older*
t	*beautifuler*

On trial $t - 1$, the subject produces an *-er* comparative. Given this response, the question that must be asked is whether the subsequent production of *beautifuler* on trial t reflects the current state of the subject's grammar (i.e., the subject's linguistic competence), or whether it is instead due to perseveration of the *-er* pattern of inflection from the previous trial. Past researchers have assumed the former interpretation (Clahsen & Temple, 2003; Graziano-King & Cairns, 2005). The latter possibility, however, cannot be ruled out.

Additionally, there is reason to believe that structuring the elicited production task so that production trials occur one after the other may inhibit our ability to detect *more* overuse. Note that in the work of Gathercole (1985) and Clahsen and Temple (2003), relatively low rates of *more* overuse are reported—3.6% and 5.9%, respectively. Graziano-King and Cairns (2005), however, document *more* overuse of up to 16.4% on adjectives that adults generally inflect with *-er*. One explanation for these differences may lie in Graziano-King and Cairn's (2005) periodic use of filler trials: three fillers were pseudo-randomly ordered among every eight comparative production trials. This may have been enough to break up strings of *-er* response perseveration, and allow for increased rates of *more* overuse. If this analysis is correct, it predicts that a more systematic and methodical use of filler trials will permit us to detect even higher rates of *more* overuse.

4.4 Experiment 5

The current experiment makes use of a method in which production trials that elicit comparative forms are methodically interleaved with fillers known as *washout trials* (Anderson, 2001). The function of washouts is to interfere with response perseveration from one production trial to the next: their role is to "wipe clean" any activation that results from exposure to the stimuli of production trial $t - 1$, so that the response on production trial t is more likely to reflect a subject's true competence.

In the experiment described below, I propose using a simple counting task as a washout. The choice of counting is based on the assumption that counting and language production are subsumed by different cognitive systems. I hypothesize that activation of the representations associated with counting will buy time for the activation levels associated with linguistic processing to return to baseline before the next production trial.

4.4.1 Participants

62 children and 20 adults participated in this study. All subjects were native speakers of English. Child participants were recruited from elementary schools in the San Diego area, and were divided into three groups based on age: 5-year-old, 7-year-old, and 10-year-old. Adult participants were undergraduate students recruited from the UCSD Department of Psychology's subject pool. Children received a certificate of appreciation upon completion of the study; some of their parents also elected to receive a

$15 remuneration. Adult subjects received course credit. Summary statistics for each subject group are given in Table 4.2. No age statistics were collected for the adult group. As a condition of participation, however, all subjects in the adult group had to be at least 18 years old.

Table 4.2: Experiment 5 participants

GROUP	N	MEAN AGE	AGE RANGE
5-year-old	19	5;5	3;10-5;11
7-year-old	21	7;0	6;1-8;3
10-year-old	22	10;0	8;7-12;5
Adult	20	—	≥ 18

All participants were tested in quiet rooms. Child participants were tested at three different local elementary schools—Mission Bay Montessori Academy, the Gillispie School, and La Jolla Country Day School—and at the Project in Cognitive and Neural Development (PCND) lab at UCSD. All adult participants were tested at UCSD's Language Production Lab.

4.4.2 Adjectives

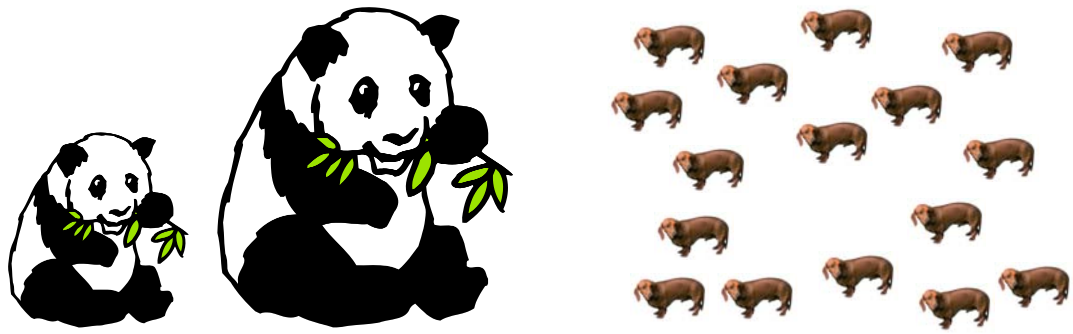
As part of another project, a set of 40 test adjectives was developed using syllabicity, frequency, and concreteness numbers assembled from the MRC Psycholinguistic Database (Coltheart, 1981). The test set varied systematically along these three dimensions such that half of the adjectives were monosyllabic and half multisyllabic, half were high frequency and half low frequency, and half were concrete while the other half were abstract. Mean log frequency differed reliably across the high and low frequency groups: 3.45 versus 0.88, $F(1, 38) = 75.70$, $p < .0001$. Likewise, concreteness ratings differed reliably across the concrete and abstract groups: 402 versus

309, $F(1, 38) = 68.25$, $p < .0001$. Across the entire set of 40 adjectives, log frequency ranged from zero to 6.74, and concreteness ranged from 198 to 473.

The effects of syllabicity, frequency and concreteness are not specifically addressed in the present set of experiments. I mention them here only to demonstrate the adjectives that our participants dealt with were quite heterogeneous in nature. All test adjectives are listed in Appendix D.

4.4.3 Materials

Production slides were created for each of the 40 test adjectives. Each production slide showed two items that differed noticeably on the attribute named by the adjective. For example, the slide for *big* consisted of two panda bears, one smaller and one larger. Additionally, 40 *counting slides* were constructed. Each counting slide showed between 12 and 17 instances of the same item, e.g. 13 lollipops, or 16 dachshunds. Example production an counting slides are given in Figure 4.1. Finally, six production slides were created for practice purposes. These slides were deliberately constructed to elicit comparatives that were highly frequent in either *-er* (*big*, *sick*, *cool*), or *more* (*comfortable*, *exciting*, *dangerous*).



Production slide Counting slide

Figure 4.1: Elicitation stimuli

Four slideshows were constructed by appending production and counting slides to the ends of empty slideshows. Slides were selected for addition to slideshows by alternately choosing one randomly-selected production slide followed by one randomly-selected counting slide until all 40 production slides and all 40 counting slides had been used. When this process was completed, the six practice production slides were added to the beginning of each show. Practice slides alternated between *-er* and *more* adjectives.

4.4.4 Method

Participants were randomly assigned to view one of the four slideshows. A single experimental session consisted of viewing 6 practice production trials, followed by 40 critical production trials interleaved with 40 counting trials. On production trials, participants viewed production slides while the experimenter pointed in turn to each of the items on the slide and uttered the frame *This X is ADJ, but this X is even* ______, e.g., *This panda is big, but this panda is even* ______. The participant would then finish the experimenter's prompt with whatever word or words they though sounded best. On counting trials, participants viewed counting slides while the experimenter prompted them to report the number of items shown on the slide.

The experimenter gave corrective feedback on practice production trials. Positive feedback was given on all other trials, regardless of the participant's response.

All child responses were videotaped for later coding and analysis. All adult responses were audiotaped.

4.4.5 Coding

Participant responses in critical production trials were coded into one of three categories: *-er*, *more*, or *other*. Unambiguous *-er* comparatives (e.g., *impossibler*, *faster*) were coded *-er*, and unambiguous *more* comparatives (e.g., *more bright*, *more elegant*) were coded *more*. All other responses were coded *other*. The *other* category included double-markings (e.g., *more easier*), instances in which participants waffled between one response and another (*nervouser…um, I mean more nervous*), and failure to produce a comparative.

4.4.6 Results

Children are said to overuse a pattern when they produce it inappropriately. In terms of comparative inflection, inappropriate use generally means either that the *-er* pattern has been misapplied to an adjective that normally takes *more* (e.g., **beautifuler,* cf. *more beautiful*), or that the *more* pattern has been misapplied to an adjective that normally takes *-er* (e.g., **more fast*, cf. *faster*). I refer to adjectives that are, by adult standards, supposed to inflect with *-er* as -ER TARGETS, and adjectives that are supposed to inflect with *more* as MORE TARGETS. Defining child misuse requires a characterization of what counts as normal adult use, i.e. what the adult targets are. In a number of past studies of comparative acquisition, researchers have defined *-er* and *more* targets intuitively, without empirically verifying what the adult behavior consists of (e.g., Clahsen & Temple, 2003; Gathercole, 1985). In the present study, however, we tested 20 adults using the exact same stimuli and procedures that were used with child participants. These adults constitute a control group against which child behavior can be compared.

4.4.6.1 Targets

Altogether, adult subjects participated in 800 critical trials. Of these, two had to be discarded due to equipment malfunction, leaving us with 798 analyzable datapoints. These data were used to identify adjectives that adults strongly favored in *more* or *-er*. For each of the 40 adjectives that adult participants were tested on, we calculated the rate of synthetic inflection—P(*er*)—and the rate of analytic inflection—P(*more*). Not surprisingly, many adjectives showed a great deal of variability. There was very little agreement, for example, concerning whether to inflect *empty*, *dense*, *narrow*, and *stupid* synthetically or analytically. Items like these exhibit a high degree of optionality, and will be considered separately in §4.4.6.3. A subset of adjectives that tended towards categorical synthetic or analytic inflection, however, could be identified. To facilitate comparisons with child inflectional behavior, I defined adjectives with adult P(*er*) values greater than 0.80 as *-er* targets, and adjectives with adult P(*more*) values greater than 0.80 as *more* targets. This procedure resulted in the 13 *-er* targets, and 12 *more* targets listed in the table below.

Table 4.3: Experiment 5 adult targets

-*ER* TARGETS		*MORE* TARGETS	
bright	neat	afraid	impossible
easy	nice	beautiful	level
fast	smooth	circular	nervous
happy	ugly	distinct	open
high	weak	elegant	orderly
long	wise	foolish	repulsive
mean			

4.4.6.2 Patterns of Overuse

In determining whether children overuse the synthetic and analytic patterns of inflection, I consider only those trials in which children inflected one of the 25 target adjectives in Table 4.3. This amounts to 2,048 datapoints, distributed roughly equally among the four age groups. For each group, I calculated mean P(*er*) and P(*more*) values in response to -*er* targets and *more* targets. Because the focus of this study is on patterns of synthetic and analytic inflection, and because *other* codes did not occur frequently or appear to vary in any interesting manner in the data, P(*er*) and P(*more*) were calculated in a way that discounted the presence of *other* codes (see 1). This did not affect the outcome of any of the analyses discussed below.

(1) P(*er*) and P(*more*) calculations

$$\mathrm{P}(er) = \frac{\mathrm{N}(er)}{\mathrm{N}(er) + \mathrm{N}(more)} \qquad \mathrm{P}(more) = \frac{\mathrm{N}(more)}{\mathrm{N}(er) + \mathrm{N}(more)}$$

Table 4.4 gives the relevant means for *more* targets and -*er* targets by age group. All rows sum to one. The proportions in bold represent possible overuse errors.

Table 4.4: Child and adult responses to *-er* and *more* targets

	P(*ER*)	P(*MORE*)
more targets		
Adult	0.01	0.99
10-year-old	**0.08**	0.92
7-year-old	**0.32**	0.68
5-year-old	**0.50**	0.50
-er targets		
Adult	0.95	0.05
10-year-old	0.89	**0.11**
7-year-old	0.75	**0.25**
5-year-old	0.75	**0.25**

Note above that the expected developmental trends occur for *more* responses to *more* targets and *-er* responses to *-er* targets. In both cases children appear to more closely approximate categorical adult inflectional behavior with age. The proportions in bold indicate possible overuse errors that explain why child behavior is not completely adultlike. I consider each set of bold responses in turn, starting with misapplications of *-er* to *more* targets.

Visual inspection of Table 4.4 indicates that adults respond with *-er* to *more* targets at a low rate—only 1% of the time. This proportion appears to be greatly inflated in some of the child groups, with 7-year-olds mistakenly inflecting *more* targets with *-er* at a rate of 32%, and 5-year-olds doing it 50% of the time. To determine whether these differences are reliable, a single-factor ANOVA was conducted using age as a between-groups variable, and P(*er*) as the dependent. Results show a significant effect of age, $F(3, 78) = 12.0$, $p < .0001$, which was explored in more detail using pairwise comparisons between each of the child groups and the adult group. One-tailed *t*-tests indicate that 5-year-olds, $t(78) = -5.29$, $p < .0001$, and 7-year-olds, $t(78) = -3.46$, $p <$

.0004, inflect *more* targets with *-er* at rates that are significantly higher than adults'. 10-year-olds and adults do not differ, $t(78) = -0.80$, $p = .21$. Means are plotted below.

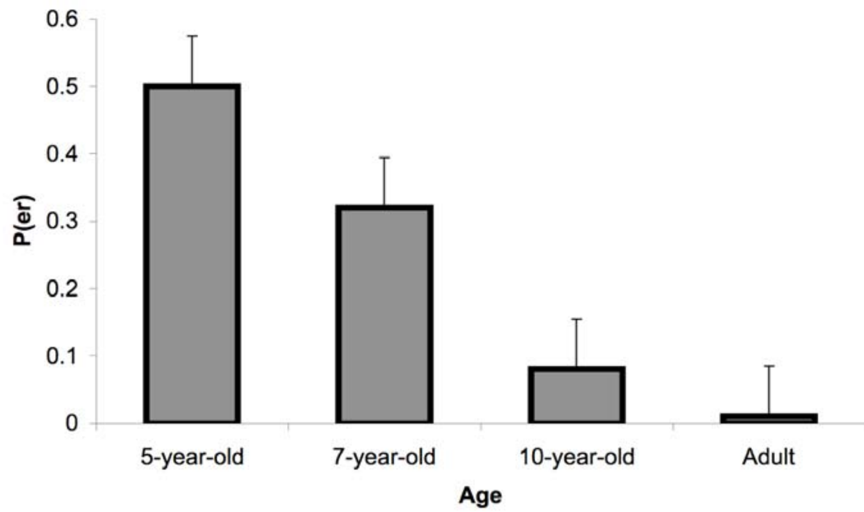


Figure 4.2: Age effects on the proportion of *-er* responses to *more* targets

This outcome replicates the patterns of *-er* overused documented in previous studies of comparative inflection (Clahsen & Temple, 2003; Gathercole, 1985; Graziano-King & Cairns, 2005), and closely matches the rates of *-er* overuse reported in Clahsen and Temple (41.6%; 2003) for 5 and 7-year-olds.

Of more interest, however, is the fact that Table 4.4 appears to show evidence of *more* overuse, which has never been convincingly documented before. Adults in Experiment 5 rarely produce *more* in reaction to *-er* targets—such responses occur at a rate of only 5%. This proportion is roughly doubled in the 10-year-old group (11%), and is multiplied by a factor of five among 7 and 5-year-olds (25%). In order to determine if these differences are reliable, I conducted a single-factor ANOVA using age as a between-subjects variable, and P(*more*) as the dependent. The results show a marginal effect of age, $F(3, 77) = 2.36$, $p = .08$. This outcome was investigated in more detail using pairwise comparisons between the adult group and each of the child groups. One-tailed *t*-tests show that 5-year-olds, $t(77) = -2.17$, $p < .02$, and 7-year-olds, $t(77) = -2.11$, $p < .02$, respond with *more* to *-er* targets significantly more often than adults do. 10-year-olds' rate of *more* usage, however, does not differ from adults on these items, $t(77) = -0.68$, $p = .25$. These findings are illustrated below.

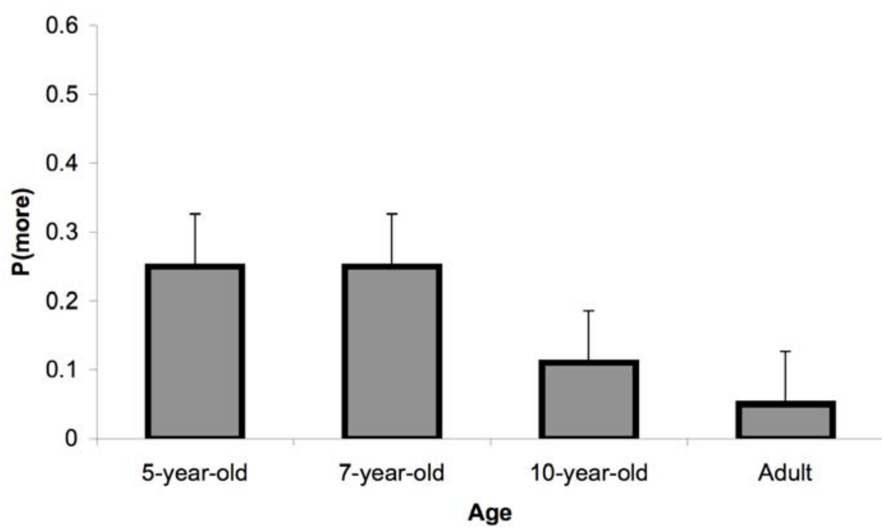


Figure 4.3: Age effects on the proportion of *more* responses to *-er* targets

These results represent the first convincing proof that children engage in overuse of the analytic pattern of inflection. They provide only weak support, however, for the type frequency-based prediction made in §4.2 of an initial stage of *-er* overgeneralization, followed later by a stage of *more* overgeneralization. According to the present data, *-er* overuse appears to peak at five years old (see Figure 4.2). But there is no concomitant *more* overuse peak at a later stage of development (see Figure 4.3). Instead, the *more* overuse peak seems to occur across both the 5 and 7-year-old groups. It is true, however, that *-er* overuse is declining at 7 years old, while *more* overuse is not. This suggests an earlier offset to the stage of *-er* overgeneralization, which may be consistent with an earlier onset. The data do not cover young enough participants, though, to say anything definitive about when children might begin to overgeneralize the different patterns on inflection.

4.4.6.3 Acquisition of Adjectives Participating in Optionality

The findings in Figure 4.3 and 4.4 establish that children overuse both the synthetic and analytic patterns of inflection. Overuse errors like these are standardly assumed to be evidence that abstract generalizations have been learned. In §4.2, I predicted that *-er* overuse would be followed by *more* overuse, based on the notion that children's input favored earlier acquisition of an abstract ADJ*er* pattern, and later acquisition of an abstract *more*+ADJ pattern. This prediction is only weakly supported in the above data, however, and may represent an idealization that is not actually present in real-world behavior. That is, it may have been naïve to believe that the different stages of overgeneralization would occur one at a time, in an orderly fashion. What may be more likely is that they overlap considerably. Significant overlap would make it difficult to uncover evidence that *-er* and *more* overgeneralization occur at distinctly different points in development. Instead, it may be that *-er* overgeneralization has an earlier onset and offset than *more* overgeneralization, but that both behaviors generally occur at the same point in time.

As a means of testing whether *-er* and *more* overgeneralization occur roughly at the same time, or whether they are significantly staggered, I considered how adjectives that inflect optionally behave over the course of acquisition. Since input from adults does not strongly indicate that these items should be inflected one way or another, it could be that they are more prone to being swept up in whatever the current developmental trend might be. For example, if *-er* and *more* overgeneralization do not significantly overlap, then optional adjectives should tend to inflect in *-er* more often among younger children, who are subject to *-er* overgeneralization, and in *more* more often among somewhat older children, who are subject to *more* overgeneralization. No such inflectional differences should be visible if *-er* and *more* overgeneralization overlap significantly.

In §4.4.6.1 I identified 25 adjectives from the test set of 40 that were (nearly) categorically inflected by adults. I was also able to find four adjectives that showed a great deal of inflectional variability. *Empty*, *dense*, *narrow* and *stupid* all have adult P(*er*) and P(*more*) values that are between 0.45 and 0.55: these adjectives were inflected by roughly half of the adult participants synthetically, and the other half analytically. I considered 328 responses to these four items across the three child groups and the adult group. P(*er*) and P(*more*) means were calculated for each group.

Table 4.5: Child and adult responses to optional adjectives

	P(*ER*)	P(*MORE*)
Adult	0.54	0.46
10-year-old	0.55	0.45
7-year-old	0.56	0.44
5-year-old	0.61	0.39

Since each row sums to one, separate ANOVA results over the means of the P(*er*) column and the P(*more*) column should be identical. I therefore conducted a single ANOVA using age as a between-subjects factor, and P(*er*) as the dependent variable. The results show no effect of age on P(*er*), $F(3, 77) = 0.12$, $p = 0.95$. I then performed a series of two-tailed *t*-tests to determine whether the slightly elevated P(*er*) value in the 5-year-old group was different from any of the other groups. This was not the case: $t(77)$ =

-0.36, $p = .72$ for the comparison with 7-year-olds, $t(77) = -0.50$, $p = .62$ for 10-year-olds, and $t(77) = -0.55$, $p = .58$ for adults. These findings are illustrated below.

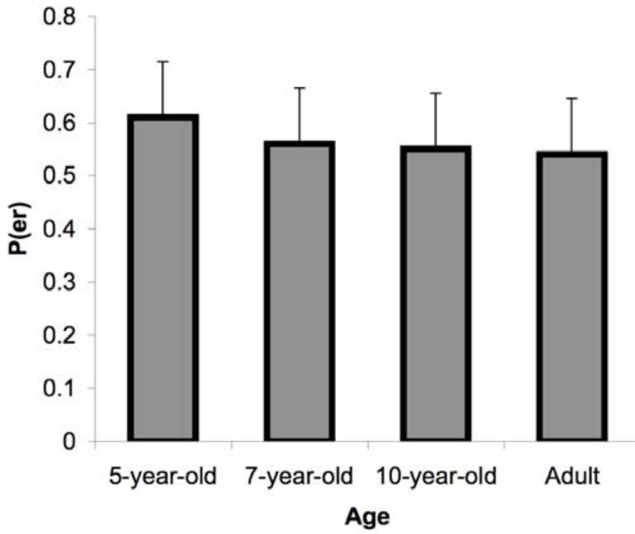


Figure 4.4: Age effects on the proportion of *-er* responses to optional adjectives

The fact that the 5-year-old rate of *-er* use is slightly elevated in comparison to the other groups hints at the possibility that the *-er* overgeneralization shown in Figure 4.2 may be affecting the way in which 5-year-olds treat adjectives that can inflect according to either the synthetic or analytic patterns. The overall outcome, however, suggests that the treatment of optional adjectives does not vary over the course of development, as might be expected if *-er* and *more* overgeneralization were developmentally staggered. These results thus seem to be consistent with the notion that *-er* overgeneralization and *more* overgeneralization overlap significantly during the course of comparative acquisition.

4.5 Experiment 6

Why did Experiment 5 find evidence in favor of *more* overgeneralization when other studies did not? Gathercole (1985), and Clahsen and Temple (2003) found that children misapplied the analytic pattern to *-er* targets relatively infrequently—only 3.6% and 5.9% of the time, respectively. These low numbers may have been dismissed as noise. Graziano-King and Cairns (2005), however, found *more* overuse at rates up to 16.4% in some adjective classes. Although they do not comment on this finding, and provide no statistical tests to verify that it constitutes a significant deviation from adult behavior, it is possible that their sparing use of washout trials somehow increased the likelihood of children responding with the analytic. If so, then Experiment 5's rigorous use of washouts may have had the same effect, albeit more strongly.

I hypothesize that the lack of systematic washout use in other studies led researchers to mistakenly underestimate the prevalence of *more* overuse. Back-to-back elicited production trials may asymmetrically encourage *-er* response perseveration (versus *more* perseveration). The synthetic pattern is supported by significantly higher token frequencies in the input. Higher frequency might conceivably lead to the establishment of a lower activation threshold for the synthetic pattern, which would make it easier to produce, especially in conditions that artificially favor the rapid production of comparative forms, one after the other. Conditions like this are prevalent in any elicited production experiment, but may have a particularly powerful effect on synthetic production in the absence of washout trials. If the synthetic pattern is subject to a higher degree of response perseveration than the analytic, then increased synthetic production may mask analytic production when no washout trials are used.

I test this hypothesis in Experiment 6 by explicitly comparing two groups of 7-year-olds: one tested without washout trials (the *no counting* condition), and one tested with systematic washout trials (the *counting* condition), as in Experiment 5. I predict increased rates of *-er* use, and depressed rates of *more* use in the no counting condition.

4.5.1 Participants

40 7-year-old native speakers of English participated in Experiment 6. I combined the 21 7-year-olds who were tested in Experiment 5 using counting trials with a new group of 19 7-year-olds who were tested using the same materials, but with all counting trials removed. Children in this new group were recruited from elementary schools in the San Diego area, and were tested in quiet rooms at either the Gillispie School, La Jolla Country Day School, or the PCND lab. All children received a certificate of appreciation upon completion of the study. Some parents also received a $15 payment for their time. Table 4.6 gives age statistics for each group.

Table 4.6: Experiment 6 participants

Group	n	Mean Age	Age Range
Counting	21	7;0	6;1-8;3
No Counting	19	6;10	6;2-8;1

4.5.2 Materials

The 19 7-year-olds in the no counting condition were tested using the same four slideshows described in §4.4.3 for Experiment 5, but with all counting slides removed.

4.5.3 Procedure and Coding

The procedure used in Experiment 6 was identical to that used in Experiment 5, except that children in the no counting condition completed production trials back-to-back, without any intervening counting trials. All critical trials were coded in the same manner described in §4.4.5.

4.5.4 Results

I decided to consider the effect of condition according to whether children were inflecting *-er* targets, or *more* targets. This allows for an assessment of whether the counting manipulation is general across different kinds of targets. Data from 1,000 production trials were considered—all trials involving the 25 adjectives in Table 4.3. The responses from one participant (25 trials total) were excluded because he produced only *other* codes which, given the method outlined in (1) of calculating P(*er*) and P(*more*), would have left us with zero-denominators. This method of calculation ensured that the P(*er*) and P(*more*) means from any participant could be summed to one, which consequently enabled us to conduct a single analysis using P(*more*) as the dependent variable, on the grounds that another analysis over P(*er*) values would simply duplicate the P(*more*) result.

To determine whether P(*more*) differed according to counting condition and target, an ANOVA was conducted using counting condition (*counting* versus *no counting*) as a between-subjects variable, and target (*-er* versus *more*) as a repeated-measures variable. The results of this analysis show, unsurprisingly, a main effect of target, $F(1, 37) = 42.50$, $p < .0001$. There was no effect of counting condition, $F(1, 37) = 1.10$, $p = .30$, and no interaction between counting condition and target, $F(1, 37) = 0.32$, p = .57. Means are plotted below.

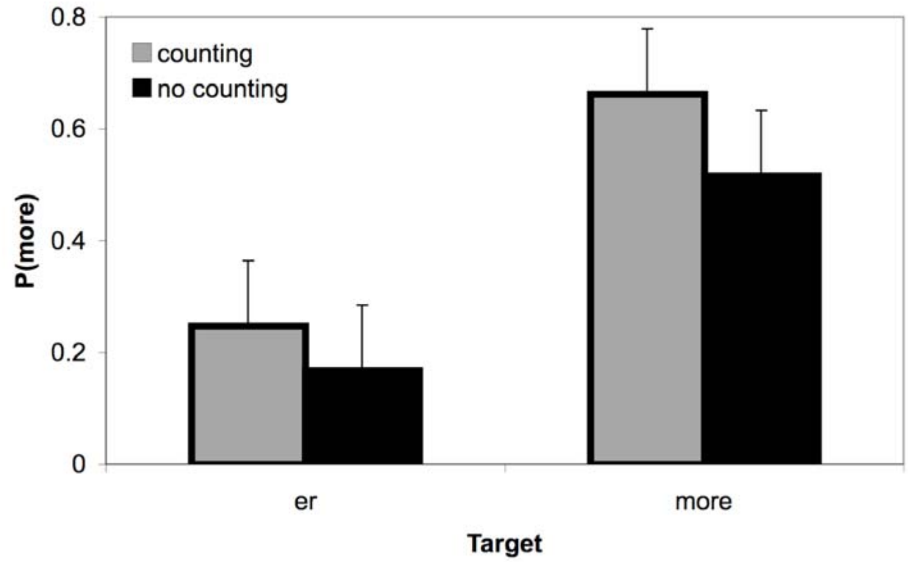


Figure 4.5: Effects of counting condition and target on P(*more*)

Although there was no significant effect of counting condition on the rate of *more* inflection, Figure 4.5 indicates a trend in the predicted direction: mean P(*more*) values are lower in the no counting group than the counting group. This trend motivated a closer examination of the data by means of pairwise comparisons of the counting manipulation within the two target groups. One-tailed *t*-tests show no effect of counting condition among -*er* targets, $t(37) = -0.96$, $p = 0.17$. There was, however, a significant effect of counting among *more* targets, $t(37) = -1.76$, $p < .05$. I interpret this result and the overall trend towards depressed *more* usage in the no counting condition as evidence that the presence or absence of washout trials in elicited production tasks can potentially affect participants' responses. The implications of this determination are addressed below.

4.6 Discussion

Experiment 5 offers the first definitive proof that comparative acquisition is characterized by not only by *-er* overuse, but also by *more* overuse. Further, these two behaviors do not seem to be segregated to distinct stages, but instead overlap considerably during development. Based on these findings, I propose a three-stage model of comparative acquisition. In Stage 1—which occurs prior to five years old—children do not form any significant generalizations over the input dataset. Linguistic behavior is conservative and item-based, as argued in a number of studies by Tomasello and colleagues (Cameron-Faulkner et al., 2003; Savage et al., 2003, 2006; Tomasello, 2000, 2003). Stage Two covers children in the five-to-eight age range, and is characterized by the induction of *general* rules of synthetic and analytic inflection, which allow *-er* and *more* to be combined with a relatively wider variety of stems than is possible in the adult grammar. Children in this stage are prone to overgeneralizing both patterns of inflection, as in Figures 4.2 and 4.3. In Stage 3, overgeneralization attenuates, presumably because children have either revised their existing synthetic and analytic inflectional rules so that they are less broad, or because the general Stage 2 rules have been retained, but are now overridden by new, narrower rules (see §5.9 for more discussion on this last possibility). In either case, the outcome is the same: by Stage 3 children's behavior is adult-like, or at least nearly so.

This progression of events makes a certain amount of intuitive sense. In Stage 1, children are collecting information about what the basic behavior should be: what kinds of wordforms should they output in order to signify the basic meaning COMP? By the onset of Stage 2, they have enough experience with English to know that there are two patterns of inflection—namely the synthetic and the analytic—but they are, simply put, not very good at using them correctly. This is akin to the acquisition of any developing skill. Becoming proficient at basketball, for example, entails first learning the basic components of the game—dribbling, passing and shooting—then refining these from their initial primitive level to a behavior that closely resembles what most people think of when they think of basketball. In comparative acquisition, this process of refinement spans the ages of five to eight. Children go into Stage 2 with basic knowledge of the comparative system in place—that there are competing *-er* and *more* patterns—but without knowing exactly how they should deploy them. This leads to overuse errors, which gradually abate as children's knowledge of the constraining factors—stem length and type of end segment (Aronoff, 1976; Frank, 1972), frequency and gradability (Graziano-King, 2003; Graziano-King & Cairns, 2005), syntactic class (Díaz, 2006), and complexity increases. Note that there is nothing in this formulation of the acquisition process that entails that sensitivity to these predictive factors will come online at the same time. The expectation is, rather, that generalizations that are more robust and reliable will be learned earlier (see §5.9 for a discussion of reliability).

The fact that children overgeneralize both the analytic and synthetic comparative patterns suggests that certain theories of competence are now less tenable. Clahsen and Temple (2003), for instance, interpret their data on comparative acquisition in terms of a dual-route model of inflection, in which regulars are generated by a symbolic rule-system, and irregulars are generated by an associative memory mechanism (Clahsen & Almazan, 1998, 2001; Pinker, 1999; Pinker & Prince, 1988). According to their analysis, overgeneralization provides the key means of determining which comparative pattern to assign to which mechanism. Since Clahsen and Temple's data suggest that only *-er* overgeneralization occurs, and because symbolic rules are touted by dual-route theorists as the only mechanism that is able to freely operate over all items in a syntactic class and thus provide that productive power needed to account for overgeneralization (Pinker & Prince, 1988), Clahsen and Temple conclude that synthetic comparatives are generated by the rule mechanism. The implication, by process of elimination, is that analytics must then be generated by the associative memory mechanism.

But this analysis becomes significantly less plausible in light of the current finding that the analytic pattern of inflection is also misused over the course of development. By Clahsen and Temple's own logic, overgeneralization of the *more* pattern indicates that analytics are also generated by the rule system. While I take no strong stand in the present work on how competence in inflectional morphology should be modeled (using rules or associative memory mechanisms), the fact that 5 to 7-year-olds produce forms like **foolisher* and **more neat* indicates that both the synthetic and analytic patterns are subject to overgeneralization, and that Clahsen and Temple (2003) therefore have no principled reason to think that that they are generated by qualitatively different systems, as suggested by dual-route theory. I conclude that a uniform analysis of comparatives is warranted, and present a candidate model of adult inflection in §5.9 that implements this desideratum. That uniformity should hold across different patterns in inflectional morphology is indicated by evidence from a plethora of other studies (Albright & Hayes, 2002; Baayen, McQueen, Dijkstra, & Schreuder, 2003; Baayen & Moscoso del Prado Martín, 2005; Plunkett & Marchman, 1991, 1993; Ramscar, 2002; Rumelhart & McClelland, 1986).

The findings of Experiment 6 imply that Experiment 5 was, at least in part, able to document *more* overgeneralization where other studies had failed to as a result of the method used. I show that *more* responses are depressed and *-er* responses are elevated when comparative forms are elicited in a way that does not make systematic use of washout trials (see Figure 4.5). This outcome has wide-ranging implications for theories of inflectional morphology that are built on data collected using the elicited production methodology. To my knowledge, no other elicited production experiment has ever attempted to control for response perseveration using washouts. This means that their data are potentially suspect because they may reflect simple carryover effects rather than true competence, as has been assumed. As noted above in my critique of Clahsen and Temple (2003), the finding that *more* overgeneralization occurs alongside *-er* overgeneralization considerably weakens their claim that the analytic and synthetic patterns behave differently over the course of development, and should therefore be modeled with different underlying mechanisms. Further research in other domains of inflectional morphology is needed in order to know whether other strong theoretical claims may likewise need to be revised.

5. Theory

The experimental results of the previous chapters suggest the need for a grammatical model of comparative inflection that is able to account for optionality. In the current chapter, I discuss a number of prominent analyses of the comparative, and show that—as currently formulated—they are all largely unable to correctly predict that a significant number of adjective stems can be inflected both analytically and synthetically. Given the empirical inadequacy of these analyses, a new way of thinking about inflection is warranted. I end this chapter by reviewing a possible solution—a model of comparative inflection that conceives of grammar as a set of lexicalized patterns, and is empirically responsive to optionality.

Theoretical analyses of the English comparative system can be taxonomized along two independent dimensions. The first dimension refers to assumptions that are made concerning class membership. On a CATEGORICAL view of class membership, adjective stems are presumed to be members of one and only one inflectional class. That is, the *-er* and *more* patterns of inflection are conceived of as being mutually exclusive: a stem's membership in one class precludes membership in the other. The categorical membership hypothesis contrasts with a position that allows class membership to be GRADIENT. According to this set of assumptions, membership in *both* inflectional classes is not ruled out: a stem could inflect with *-er* on some occasions, and with *more* on others.

The second dimension along which analyses of the English comparative can be situated concerns the division of labor within the system. Is responsibility for the generation of synthetic and analytic forms divided across more than one linguistic domain, or are synthetics and analytics treated uniformly in the same domain? According to what I will refer to as MIXED analyses, synthetics are generated lexically, in the morphology, while analytics are handled in the syntax. This is in opposition to UNIFORM treatments, in which a single module (either morphology or syntax) is responsible for the generation of both synthetics and analytics.

These different dimensions of comparison can be crossed to give a four-celled paradigm in which existing theoretical proposals can be placed, as in Table 5.1.

Table 5.1: Taxonomy of comparative analyses

CATEGORICAL		GRADIENT	
MIXED	UNIFORM	MIXED	UNIFORM
Poser (1992)	Embick (2007) Stump (2001)		Bochner (1993)

Poser's (1992) analysis, for instance, is classified as *categorical-mixed*, while analyses by Embick (2007) and Stump (2001) fall into the *categorical-uniform* group. Note that the cell for *gradient-mixed* proposals is empty. While there is no *a priori* reason why such an approach cannot be worked out, I know of no existing analyses that falls under this heading. Finally, Bochner's (1993) work exists as an exemplar of the class of *gradient-uniform* accounts of comparative inflection. I refer to these categories throughout the discussion that follows.

5.1 Blocking and Optionality

In this section I situate the phenomena known as *blocking* and *optionality* with respect to categorical and gradient morphological analyses, and provide a simple formalization of each. The term blocking is often associated with a categorical view of class membership. Aronoff (1976) gives a simple example of blocking in his discussion of how the adjective *good* is inflected for comparison. *Good* is a high-frequency, monosyllabic stem. Consequently, we might expect it to be a member of the *-er* inflectional class, as these features tend to correlate with *-er* suffixation (Graziano-King, 2003). Membership in this class would entail that *good* be inflected as **gooder*. Instead, however, *good* belongs to a single-member class that forms the comparative by combining the suppletive stem *bett-* with the *-er* suffix to produce *better*. Aronoff claims that *better* occurs while **gooder* does not because lexically-listed *better* blocks the normal comparative formation process that would otherwise result in **gooder*. Further, the blocking of **gooder* by *better* obtains on every production: *better* is *always* the form used to express the meaning GOOD in the comparative: **gooder* is *never* used. This is equivalent to saying that *good* only has class membership in its own, suppletive inflectional class; it cannot also be a member of the *-er* class. A blocking analysis thus presupposes categorical membership.

Optionality, on the other hand, is associated with a gradient view of class membership. English verbs such as *dive*, for example, seem to participate in multiple patterns of past tense inflection: *dived* and *dove* are both legal possibilities for some speakers (Pinker, 1999). The same semantic content—DIVE plus the morphosyntactic feature PAST—thus appears to be realized in multiple ways. On an Aronoff-style blocking analysis, this kind of variation would not be predicted. Lexical listing of *dove* would eliminate the possibility of *dived* ever being generated. Instead, however, the class membership of *dive* appears to be gradient in nature: it is a member of both the regular *-ed* inflectional class, and the /aɪ/ ~ /oʊ/ ablaut class.

5.2 A Uniform Account of Blocking and Optionality

Blocking and optionality overlap in that they both describe situations in which a meaning has competing modes of expression. They differ in terms of whether they allow for the possibility that multiple expressions might actually be used. In the blocking relationship, competing expressions are always mutually exclusive. Under optionality, however, no competitor is categorically precluded from appearing. Although previous proposals have failed to recognized it, there is an intuitive sense in which the apparently distinct relationships described by blocking and optionality are simply different facets of a single, unified phenomenon.

In order to capture the similarities and differences between blocking and optionality, I adopt a straightforward formalization consisting of three very simple equations. I assume a mapping in which a meaning—represented by a lexeme L and its associated morphosyntactic feature set σ—can be realized by any member of an arbitrarily large set of expressions, $\{e_1, e_2 \ldots e_n\}$. This assumes a variant of the Sausurrean approach to signs whereby content, designated with ⟨L, σ⟩, is realized by some specified form or forms. Under both blocking and optionality, the probability of use assigned to each of $\{e_1, e_2 \ldots e_n\}$ is constrained by (1).

(1) $P(e_1) + P(e_2) + \ldots + P(e_n) = 1$

This formulation serves to highlight what is shared between blocking and optionality: an increase in the probability that e_n will be produced leads to a subsequent decrease in the probability of use of at least one of its competitors. The basic distinction between blocking and optionality is made explicit in (2) and (3).

(2) Blocking Hypothesis for Competing Realizations

$P(e_n) = \{0, 1\}$

(3) Optionality Hypothesis for Competing Realizations

$P(e_n) = [0, 1]$

According to what I refer to as the Blocking Hypothesis for Competing Realizations, the values that $P(e_n)$ can take are denoted by the set of zero and one. In conjunction with (1), this entails that if $P(e_n) = 1$, then no other expression will ever surface. Likewise, if $P(e_n) = 0$, then one of e_n's competitors will necessarily take over—e.g., $P(e_{n-1})$ might be set to 1 in order to comply with the restriction in (1).

This categorical outcome in production probabilities differs from what is described in (3), the Optionality Hypothesis for Competing Realizations. In this case, the value of $P(e_n)$ is not restricted to zero or one. Instead, $P(e_n)$ can take on any real-number value ranging from zero to one, inclusive. Given, for example, a meaning with two competing expressions—e_1 and e_2—this entails that if $P(e_1)$ were set to 0.6, then $P(e_2)$ would consequently equal 0.4. Note that while the Optionality Hypothesis does allow $P(e_n)$ to take on values other than zero and one, it does not prohibit the kinds of categorical outcomes that characterize the Blocking Hypothesis. In terms of the range of mappings available under the Optionality Hypothesis and the Blocking Hypothesis then, it should be apparent that all mappings described by the Blocking Hypothesis can also be described by the Optionality Hypothesis, but not vice versa. The insight provided by the formulations in (1-3) is thus that a gradient approach to inflectional class membership subsumes categorical treatments.

This line of inquiry provides the basis for a uniform treatment of blocking and optionality. Rather than approaching blocking effects as resulting from specific blocking principles, categorical outcomes can now be analyzed as highly probable mappings within a flexible system that is able to account for a wide range of empirical facts—from robust judgments about forms like **beautifuler* and **more big*, to indecisiveness about whether to inflect *likely* as *likelier*, or *more likely*.

5.3 Manifestations of Blocking and Optionality

Nothing in the definitions provided in (1-3) constrains the actual surface form of e_n. This is important because blocking and optionality do not appear to be sensitive to the way in which e_n is realized: neither seem to care whether the expressions they regulate are synthetic, or analytic. Aronoff's (1976) analysis of *better* and **gooder*, for example, illustrates that synthetic forms can block other synthetics. This is traditionally the way in which blocking has been conceptualized. More recently, however, researchers have noted that the blocking relationship seems to extend to competitions between synthetic and analytic forms. Stump (2002), for example, shows that the analytic realization of a nominal's number and class in numeral-classifier constructions in Malto is precluded in exactly those instances in which a synthetic form is already available.

(4) Blocking between synthetic and analytic expressions (Stump, 2002)

a. tīni jen maler
three CLASSIFIER men
'three men'

b. *eike jen maleh
one CLASSIFIER man
'one man'

c. ort maleh
CLASSIFIER.one man
'one man'

Numeral-classifier constructions usually consist of a numeral followed by a classifier (*jen* is the classifier for humans in Malto) followed by a noun. This is the pattern shown in (4a). For number amounts less than three, however, the expected analytic numeral-classifier expression is unavailable (see 4b), and is instead replaced by a suppletive synthetic expression, as in (4c). This phenomenon has an analysis that closely parallels that given in the *better*-**gooder* case: an expression that is expected on independent grounds fails to appear because it is blocked by a competitor.

Synthetic-analytic blocking effects like those in (4) have been reported in a number of other languages as well. Mari and Udmurt verbal predicates (Ackerman & Stump, 2004), Tundra Nenets nominal declension (Salminen, 1997), Irish conditionals (Andrews, 1990), the Basque progressive (Poser, 1992) and, of course, the English comparative (Embick, 2007; Poser, 1992; Stump, 2001) are all domains in which synthetic forms appear to block analytic competitors. The reverse relationship also arguably obtains. Sadler and Spencer's (2000) analysis of the Latin perfective passive suggests that analytic forms can block synthetics (but see Stump, 2002 and Kiparsky, 2004 for a critique of this position). And Brassil (2001) argues that some analytic constructions are able to block other analytic constructions. Together, these examples indicate that blocking is not restricted to one-word forms. Instead, the available evidence demonstrates that blocking is indifferent to the form of the expressions that it regulates.

Optionality shows a similar indifference to surface realization. Chapter 1 introduced a wide range of examples of morphological optionality. In some of these, synthetic forms compete with other synthetics, as in English past tense realization (Pinker, 1999) and free prefix ordering in Chintang (Bickel et al., 2007). In others, multiple analytic forms compete (e.g., the Bulgarian negative future perfect; Manova, 2006), or synthetic-analytic competition occurs (as in the Greek plural genitive or English comparative; Aronoff, 1976; Frank, 1972; Graziano-King, 2003; Sims, 2005). Cumulatively, these observations suggest that the formulation in (1-3) is essentially correct in declining to constrain the surface realization of e_n. Whether the competition between competing expressions involves one-word or multi-word forms is orthogonal to the difference between blocking and optionality. Of primary importance is the connection between blocking and a categorical view of inflectional class membership, and between optionality and a gradient view of class membership.

5.4 A Categorical-Mixed Comparative Model

Above, I note that blocking and optionality occur among competitors that are uniformly synthetic, uniformly analytic, or divided between synthetic and analytic realizations. Assuming that one-word forms are morphologically-generated, then synthetic-synthetic interactions—such as the blocking of **gooder* by *better*—are entirely lexical in nature. It is unclear, however, whether the same can be said of synthetic-analytic blocking. On the assumption that one-word forms are generated in the lexicon via morphological processes, while multi-word forms are generated in the syntax, it would appear that synthetic-analytic blocking spans the two grammatical components. Syntactically-generated **more big*, for example, would seem to be blocked by lexical *bigger*.

Poser (1992) adopts precisely this view of the division of labor between morphology and syntax. This feature of his analysis situates it squarely as an exemplar of the class of *mixed* models introduced at the beginning of this chapter. According to Poser, English comparatives are by default formed analytically in the syntax. Analytic expression is blocked, however, when lexically-generated synthetic forms are available. The availability of such forms is phonologically-conditioned based on the number of syllables in the stem. Generally, "adjectives with mono- and di-syllabic stems are perfect [i.e., capable of forming the comparative synthetically], while lexical forms derived from adjectives with longer stems are unacceptable" (p. 121). This condition allows a short stem like *big* to have a synthetic form *bigger*, which subsequently blocks **more big*. It also accounts for the fact that longer stems, such as *dangerous*, tend to form the comparative analytically. Since *dangerous* is trisyllabic, it does not have a synthetic realization available (**dangerouser*), no blocking occurs, and syntactically-generated *more dangerous* surfaces. In summarizing his analysis, Poser (1992: 121) writes "The crucial observation is, then, that whatever the nature of the principles governing the well-formedness of lexical comparatives…the acceptability of the periphrastic [i.e., analytic] forms is inversely related to that of the lexical forms…. As far as I can see, the only plausible explanation for this is that the periphrastic [analytic] forms are blocked by the lexical forms."

This conclusion follows from the assumption that one-word versus multi-word surface expression leads to a transparent assignment of forms to the lexical versus the syntactic components of the grammar: if a form is synthetic, then it is located in the lexical component, whereas if it is analytic it must be located in the syntax component. The apparent naturalness of this assumption aside, in order for Poser (1992) to establish that the comparative domain is primarily characterized by blocking effects, it is first necessary to show that adjective stems are categorically inflected either as synthetics or analytics—i.e., that optionality does not occur. But the results described in the previous chapters forcefully demonstrate that this is an unwarranted idealization. Instead, optionality seems to be systematic and relatively widespread. Poser, in fact, provides tacit support for this position when, in a footnote (p. 121), he acknowledges that monosyllabic stems appear to be subject to consistent analytic inflection in metalinguistic comparative constructions (which will be treated in detail in §5.5). Given this admission, Poser's (1992) claim that English comparative inflection counts as an example of blocking is, accordingly, weakened. At best, the claim should be tempered to reflect that certain adjective stems behave according to blocking principles, while others do not. Alternatively, a more straightforward analysis would treat all adjectives as falling within the optionality hypothesis. Given the twin observations that optionality is required to account for cases such as *angrier* ~ *more angry* and *prouder* ~ *more proud*, and that optionality subsumes blocking behavior according the definitions in (1-3), the simplest way to provide coverage of adjectives that behave categorically as well as gradiently is clearly to adopt the optionality hypothesis.

Poser's (1992) claim that analytic comparatives are syntactic is also largely unmotivated. As evidence in favor of the presumably syntactic status of analytic comparatives, he cites the examples below.

(5) more [curious and inquisitive]

(6) The situation is more, I suppose the term is delicate, than I had thought.

(5) illustrates that *more* can be used to modify arbitrary conjunctions of adjectives, while (6) shows that *more* can be syntactically separated from the adjective that it modifies by appositive material. Both of these phenomena indicate that the pieces of the analytic comparative construction are syntactically independent. However, syntactic independence does not necessarily demonstrate that analytic comparatives are syntactic rather than lexical entities.

While it is accurate to observe that many researchers in both lexicalist and nonlexicalist frameworks assume the syntactic status of multi-word expressions, considerable evidence from traditional as well as contemporary morphological theory challenges this assumption. A large number of analyses covering a wide range of inflectional and derivational phenomena suggest that analytic combinations (periphrases) are the product of morphological operations—i.e., that apparent syntactic combinations are actually multi-word morphological expressions (Ackerman & LeSourd, 1997; Andrews, 1990; Sadler & Spencer, 2000; Stewart & Stump, 2007; Stump, 2002). This principle is articulated by Ackerman and Stump (2004) as the Periphrastic Realization Hypothesis, which states that inflectional rules are capable of defining both synthetic and analytic forms. Further, an analytic pattern is lexical if—among other defining properties—its distribution is *featurally intersective*. By this, Ackerman and Stump refer to cases in which a pattern fails to exclusively align with the expression of any particular morphosyntactic feature. It is not the case, for example, that COMP can only be realized analytically: synthetic expression is also a possibility. The fact that COMP can be realized synthetically demonstrates that morphology has access to COMP and is able to use it in the word formation process. Under this circumstance, analyzing analytic comparatives as syntactic leads to a duplication of labor—both grammatical components would be building expressions with the exact same meaning. The simpler alternative is to allow morphology to create analytic forms when there is independent evidence from synthesis that the proper morphosyntactic features are already available. This kind of reasoning goes beyond the arguably naïve approach of declaring *all* multi-word expressions syntactic, and offers a principled foundation for the lexical analysis of some analytic expressions, like English *more* comparatives.

The examples in (5) and (6) that Poser (1992) offers as evidence in favor of the syntactic status of analytic comparatives are actually entirely consistent with the kind of morphological analysis just described. Ackerman and Stump (2004: 122) propose that the insertion of periphrases into syntactic structure is regulated by the Periphrastic Realization Principle.

(7) Periphrastic Realization Principle

> Where the realization of w_1 w_2 of ⟨L σ⟩ is periphrastic [analytic] and w_1 and w_2 belong to the respective categories X and Y, w_1 and w_2 may be inserted as the heads of the respective phrases XP and YP.

As Ackerman and Stump note, this formulation "makes no claims about the surface constituency relations among elements of periphrastic constructions" (p. 122). It thus follows that lexically-generated analytic comparatives can be inserted into many different kinds of syntactic structures. In many cases, the structure will maintain linear adjacency between *more* and the adjective stem. It is also possible, however, that the different pieces of a morphological analytic expression will be separated in the syntax, as in (5) and (6).

I conclude this section by rejecting two central tenets of Poser's (1992) analysis. First, it is not at all apparent that inflection in the comparative domain is best modeled categorically, using blocking principles. The available evidence instead suggests that a gradient model that includes the possibility of optionality provides the best fit to the data. Second, the fact that analytic comparatives are composed of more than one word, and that these words can be linearly separated in the syntax, does not necessarily indicate their syntactic status. Borrowing from principles developed by word-based realizational morphologists,[9] I outline an alternative analysis in which some multi-word expressions are generated in the lexical/morphological component of the grammar.

5.5 A Syntactic Categorical-Uniform Comparative Model

I now turn to a recent analysis of the English comparative by Embick (2007) that is formulated within Distributed Morphology, a syntactocentric theory of word formation (DM; Embick & Noyer, 2007; Halle & Marantz, 1994). Like all theoretical frameworks, DM is built upon number of assumptions. One of the most provocative—since it departs

---

[9] See §5.8 for an overview of word-based morphology.

from both pedagogical tradition and time-honored morphological insights—is the claim that word formation is explained and constrained by the types of principles and representations favored in Chomskyan syntactic analysis, rather than being the product of distinctly morphological operations. This principle is referred to by Halle and Marantz (1994) as Syntactic Hierarchical Structure All the Way Down (henceforth, SHSAWD), and leads to the uniform nature of Embick's (2007) analysis: since there is no autonomous lexical component, all comparatives—synthetics included—are syntactically-generated.

Two further assumptions made in Embick's analysis are outlined in Culicover and Jackendoff's (2005) critique of mainstream generative grammar. These are known as Interface Uniformity (IU) and Structural Uniformity (SU; pp. 6-7). According to IU, semantic structure maps transparently and directly onto syntactic structure. This entails that the same meaning will always map onto the same structure, and that different syntactic structures will have different meanings. The assumption of IU without SU is potentially problematic for the analysis of comparatives. On the face of it, synthetics and analytics have different structures: the comparative marker is suffixed in synthetics, whereas it occurs before the stem in analytic comparatives. This structural difference should, according to IU, signify a difference in meaning. But this undesirable outcome is avoided by further assuming that synthetics and analytics must, at some level of representation, have the same structure—i.e., that they obey structural uniformity. IU and SU thus jointly motivate additional assumptions: a derivational theory of word-formation, and movement operations that allow synthetic comparatives to occur as transformations from a shared underlying analytic structure that interfaces with semantics. It is, of course, difficult to know whether these additional postulations provide any actual insight into the grammar of word formation, or whether they are simply auxiliary assumptions necessitated by the erroneous adoption of IU and SU. In any case, the collective effect of SHSAWD, IU and SU is to ensure a central role for syntactic representation in explanations of morphological phenomena. SHSAWD does this by eliminating morphology and the lexicon as a distinct grammatical components. IU and SU accomplish the same by positing a semantics in which meaning is straightforwardly assessed on the basis of syntactic trees—i.e., where semantic structure and syntactic structure are isomorphic at the interface.

In Embick's (2007) analysis, analytics occur as the default comparative realization, and synthetics only surface when certain phonological and syntactic conditions (discussed below) are met. All comparatives share an underlyingly analytic syntactic structure in which the degree marker (Deg) heads a Degree Phrase, and takes an Adjective Phrase (AP) as its complement.

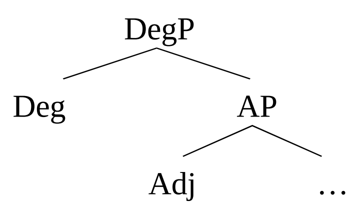


Figure 5.1: Comparative structure in (Embick, 2007)

During the course of an analytic derivation, abstract morphemes such as Deg and Adj are specified in the structure in Figure 5.1, and take on the correct semantic interpretation as a result of their phrase-structural status as heads of their respective phrases. Both morphemes are phonetically realized using items drawn from the Vocabulary—a non-generative repository of forms in DM. For an analytic comparative, this means that Deg and Adj will be realized by *more* and an adjective stem, and that the proper linear order and syntactic scope relations fall out automatically as a result of the way in which the tree is composed.

Synthetic realization is somewhat more complex. Because synthetics have the same meaning as analytics, and because meaning is read directly off of syntactic trees, synthetic and analytic comparatives must share the same structural representation at some stage of the relevant derivation. This is accomplished by appealing to Local Dislocation, an operation that moves Deg from its underlying (analytic) position in Figure 5.1, to a surface (synthetic) position in which it suffixes to the adjective stem to create a word of the form ADJ*er*. Whether or not Local Dislocation applies depends on two conditions being met: the adjective stem must be of the right phonological shape (the *prosodic condition*), and the stem must be in the correct tree-configural relationship with Deg (the *adjacency condition*). I treat each of these conditions in turn below.

Embick defines the prosodic condition very loosely, saying only that the derivation that creates synthetics applies exclusively to "short" adjectives (p. 9). As reported elsewhere in this work (e.g., §1.3), it is commonly recognized that inflection with *-er* correlates with stem length such that it becomes increasingly uncommon to see synthesis as the number of stem syllables grows past one (Aronoff, 1976; Graziano-King & Cairns, 2005; Mondorf, 2003; Poser, 1992). The prosodic condition thus represents a feature that is common to all comparative analyses. It is used here, however, to enforce a categorical division between short and long adjectives. Synthesis is blocked for longer adjective stems because Local Dislocation can only apply when the prosodic condition is met, and it is only met when the stem is short enough. Note that a categorical division between *-er* and *more* comparatives is achieved here in a manner that differs somewhat from blocking as it is described elsewhere. Blocking effects typically occur when the existence of one form precludes the use of a form that might otherwise be expected: words block words (Aronoff, 1976), or synthetics block analytics (Poser, 1992). According to Embick's analysis, however, it is not that one form blocks another, but that certain outcomes are *derivationally* precluded because the application of Local Dislocation is conditioned on stem length.

Embick (2007) differs strikingly from other comparative analyses in the explanatory power that he attributes to syntactic principles. The second condition that he places on the application of Local Dislocation, for example—the adjacency condition—blocks synthesis when Deg and Adj are not in the correct tree-configural relationship. The adjacency condition is motivated, according to Embick, based on data from two types of phenomena: adverb interposition and metalinguistic comparison. For brevity, I focus here on addressing Embick's claims concerning metalinguistic comparison, and simply note that adverb interposition is straightforwardly dealt with under standard lexicalist assumptions.[10] Examples of metalinguistic comparative constructions (Bresnan, 1973; McCawley, 1988; Poser, 1992) are given below.

(8) Metalinguistic comparison (Embick, 2007)

a. John is more lazy than stupid.

b. *John is lazier than stupid.

c. It is more hot than humid.

d. *It is hotter than humid.

---

[10] See Boyd and Ackerman (2007) for detailed discussion.

Normally, short adjectives such as *lazy* and *hot* would be inflected as *lazier* and *hotter*. As (8b) and (8d) attest, however, synthetic inflection appears to be dispreferred in metalinguistic comparison. Embick analyzes these cases as resulting from a subtle semantic distinction. In normal comparatives, such as those in (9), two entities are being compared on a single dimension named by an adjective:

(9) a. John is lazier than Bill.

b. It is hotter today than it was yesterday.

In (9a), John and Bill are situated with respect to one another on a scale of laziness. Similarly, in (9b), today is compared with yesterday along the dimension of hotness. The examples in (8) differ from those in (9) in that the metalinguistic comparatives can be paraphrased as saying that it is *more appropriate* to label John as lazy than stupid (8a), and that it is *more appropriate* to label today as hot than it is to label it as humid (8c). That is, in metalinguistic comparison, the dimension along which the comparison is being made is actually one of *appropriateness*, and has very little to do with the adjectives that are named.

Since, according to Interface Uniformity, all semantic distinctions are marked in the syntax, the proposed difference in meaning between normal comparatives and metalinguistic comparatives must have a syntactic locus. This is outlined as follows.

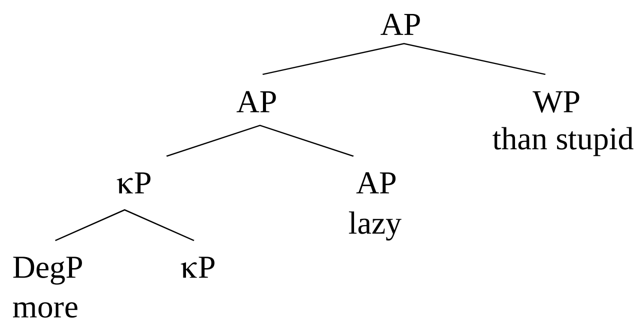


Figure 5.2: The structure of metalinguistic comparatives (Embick, 2007)

The meaning element that leads to a different interpretation for metalinguistic comparatives versus normal comparatives is located in a silent adverbial that Embick labels κP. Crucially, κP is located between DegP and the adjective stem. This has the effect of disrupting the linear adjacency relation that holds between Deg and Adj in the structure of normal comparatives, as in Figure 5.1. Because Deg and the adjective stem are not contiguous in metalinguistic comparatives, the adjacency condition is not met, Local Dislocation cannot not apply, and synthetic realization is barred, even for short adjectives that might otherwise be suffixed, e.g., *lazy* and *hot*. This accounts for the judgments cited in (8).

A generous interpretation of this proposal suggests that the presence of κP predicts the departures from synthetic expression that are observed in metalinguistic comparison. Short adjectives are not always inflected with -*er*. Instead, short stems are sometimes inflected with *more*, depending on the type of comparative construction they occur in. This arguably represents a constrained and specific form of alternative encoding that complements Embick's otherwise categorical system of inflection. The predictions of a model including both the prosodic and the adjacency conditions are (i) that "short" adjectives in normal comparatives will never be inflected analytically, and

(ii) that synthetics will never be allowed in metalinguistic comparison (see 8b and 8d). This first prediction is not borne out in experimental studies of comparative inflection. As noted in §1.3, Graziano-King (2003) found that speakers routinely favored the analytic variant over the synthetic for low-frequency monosyllabic stems such as *apt* (**apter*, cf. *more apt*). Moreover, the data discussed in Chapters 3 and 4 demonstrate that monosyllabic adjective stems show a great deal of systematic inflectional variability. I evaluate the prediction in (ii) below.

According to Embick's (2007) analysis, since synthetics are blocked in metalinguistic comparison due to the presence of an intervening κP, short adjectives that would otherwise take *-er* will always be inflected with *more*. The data, however, are somewhat more nuanced than Embick's proposal acknowledges.[11] While perhaps not perfectly acceptable, the sentences in (10) appear to be significant improvements over (8b).

(10) a. John is far lazier than stupid.

b. John is lazier than he is stupid.

c. John is far lazier than he is stupid.

Sentence acceptability seems to be modulated by the presence of an adverbial modifier before the initial adjective of the comparative clause—as in (10a)—and also by the introduction of (perhaps elided) predicate structure—as in (10b). The combination of

---

[11] To be fair, a truly data-driven critique of Embick's (2007) claims regarding metalinguistic comparatives would require experimental work of the type done on ordinary comparative formation in Chapters 2, 3 and 4 of this thesis. For the moment, I must be content to evaluate Embick's remarks using data collected using the same methods he does—i.e., through the unsystematic elicitation of native speaker judgments. While not ideal, this technique at least provides data for further inquiry.

these two elements in (10c) leads to a sentence that sounds nearly as acceptable as the version in (8a).

These examples are quite problematic for Embick's analysis because they are all metalinguistic in nature. The same paraphrase that applies to (8a)—that it is more appropriate to call John lazy than stupid—also applies to every sentence in (10). According to Embick then, these sentences should all have the structure given in Figure 5.2: a κP should intervene between DegP and the adjective stem. If they all contain a κP, then the adjacency condition is not met, Local Dislocation cannot apply, and the synthetic forms in (10) should all be robustly ungrammatical. This is obviously not the case, however. Instead, we have a range of metalinguistic comparatives in *-er* that, contrary to the predictions of Embick's analysis, are not precluded.

It is, of course, possible to deploy Embick's κP device wherever it yields the correct distribution of forms. Anytime a short adjective failed to inflect in *-er*, we could simply posit a structural difference—a phonetically unrealized κP—that would serve to block the derivation of a synthetic form. Clearly, however, this no longer retains the principled explanatory power it was designed to achieve. Equally worrisome is the fact that the data from metalinguistic comparison were offered as independent evidence in support of the assumption that the distribution of comparative forms is subject to a predominantly syntactic analysis. If closer examination of metalinguistic comparison (as exemplified by the examples in 10) suggests that it does not offer such support, then the proposed κP node simply accounts for the data that it set out to address—that is, it simply redescribes a subset of the data using favored assumptions.

A possible alternative approach would be to abandon the notion that the distribution of analytic and synthetic forms is explicable on the basis of purely syntactic assumptions. A more careful examination of the data (see Chapter 3, for example) indicates that the form that an adjective takes in the comparative is determined by a multitude of factors: phonological, semantic, and syntactic. The hypothesis that linguistic objects need not have syntactic origins, and that grammatical phenomena can have multifactorial causes is, in fact, exactly what is argued in some recent critiques of mainstream generative grammar (Culicover & Jackendoff, 2005; Jackendoff, 2003). I will attempt to capitalize on this new way of thinking about language in the model of comparative formation that I propose in §5.9.

5.6 A Morphological Categorical-Uniform Comparative Model

In my critique of Poser's (1992) mixed model in §5.4, I argued that multi-word expressions need not be generated syntactically. The fact, for example, that the pieces that make up an analytic comparative can be separated by intervening words does not convincingly demonstrate a syntactic genesis. Instead, under Ackerman and Stump's (2004) Periphrastic Realization Hypothesis, multi-word expressions can be lexically-generated if the right conditions are met. In the case of English comparatives, since the existence of synthetic forms shows that the morphological component has access to the grammatical feature COMP, and is able to create wordforms that bear COMP, analyzing analytic comparatives as syntactic creations constitutes a duplication of labor. Positing that some comparatives are generated in the morphology while others are generated in the syntax misses the fact that these forms are semantically identical, and differ trivially based on form.

In §5.5, we saw that it is possible to construct a comparative analysis that treats both synthetics and analytics in the same manner—i.e., syntactically. But Embick's (2007) proposal departs sharply from longstanding traditions in morphological analysis by making a number of assumptions whose collective effect is to marginalize the explanatory and generative power of non-syntactic grammatical components (Culicover & Jackendoff, 2005; Jackendoff, 2003). Embick's approach is representative of the ideological leanings of many adherents to mainstream generative grammar, but fails to provide adequate empirical coverage of optionality, and particularly of the range of forms that are available in metalinguistic comparative constructions.

In the present section, I review a proposal by Stump (2001) that attempts to do away with the duplication of labor present in Poser (1992) by taking a very different tack than Embick (2007). Rather than claim that all comparatives have a syntactic genesis, Stump (2001) adopts the Periphrastic Realization Hypothesis, and treats synthetic and analytic forms uniformly in the lexicon.[12] Stump formulates his proposal in the language of exponence-based realizational morphology, which rejects the notion that words are composed of meaning-bearing primitive forms known as *morphemes*. According to Stump's approach, all paradigms are described by a paradigm function that takes a root-morphosyntactic feature set pair as input, and outputs the entire paradigm that the root

---

[12] The analysis presented in Stump (2001) does not address analytic inflection. What I summarize here is a modification that is consistent with Stump's later adoption of the Periphrastic Realization Hypothesis (Ackerman & Stump, 2004; Stewart & Stump, 2007; Stump, 2002).

participates in. The input pair is written as ⟨X, σ⟩, where X is a root, and σ is a well-formed morphosyntactic feature set. The paradigm function associates the input pair ⟨X, σ⟩ with specific surface realizations by calling a series of realizational rules (RRs). Realizational rules are organized into ordered blocks, with the rules in each block in competition with each other according to the principle of Pāṇinian determinism. This amounts to the strong claim that grammars are organized in such a manner that blocking effects simply follow without the need to specify a separate blocking principle. Among the stipulated set of rules that can apply in any given rule block, only the narrowest rule actually is applied. For comparative formation, Pāṇinian determinism ensures that *-er* suffixation (which only applies to short adjectives) blocks inflection with *more* (which applies to any adjective, and is therefore less narrow). The realizational rules in (11) outline Stump's full analysis.

(11) Summary of Stump's (2001) comparative analysis

Block 1	Block 2
(a) $RR_{\{ADJ\}}$(⟨X, σ⟩) = $_{def}$ ⟨Y, σ⟩, where Y is X's bare stem.	(d) $RR_{\{ADJ\}}$(⟨X, σ⟩) = $_{def}$ ⟨*more* X, σ⟩
(b) $RR_{\{GOOD\}}$(⟨X, σ⟩) = $_{def}$ ⟨*bett*, σ⟩	(e) $RR_{\{SHORT\ ADJ\}}$(⟨X, σ⟩) = $_{def}$ ⟨X*er*, σ⟩ [defined only if X meets certain prosodic conditions, which ordinarily entail that X has fewer than three syllables]
(c) $RR_{\{FAR\}}$(⟨X, σ⟩) = $_{def}$ ⟨*furth*, σ⟩	

(f) $RR_{\{BAD\}}$ (⟨X, σ⟩) = $_{def}$ ⟨*worse*, σ⟩

The analysis is divided into two rule blocks. Each block has responsibility for building part of a comparative form. Block 1, for example, handles stem selection. Note the specifications on rule application listed in the subscripted curly braces after each rule's RR notation. These indicate that (11a) is the default rule in Block 1 because it applies to all adjectives, whereas the narrower (11b) and (11c) only apply to GOOD and FAR. For most inputs to Block 1, the default rule—(11a)—will select the single stem Y associated with X. Given the input pairs ⟨*big*, {COMP}⟩ and ⟨*dangerous*, {COMP}⟩, for example, *big* and *dangerous* will be selected. Suppletive items such as GOOD and FAR, however, have more than one stem. In these cases, Pāṇini's principle applies, and the narrower rules in (11b) and (11c) block (11a), and output *bett* and *furth*, respectively.

Block 2 operates over the stems generated in Block 1 by adding the exponents -*er*, or *more*. In the default case described in rule (11d), *more* is added to a stem to create an analytic comparative of the form *more* X. This is what occurs for the input stem *dangerous* (→ *more dangerous*). As in Block 1 though, the default can be overridden. The narrower rule in (11e) applies only to short adjectives. The result is that, when a short stem is input to Block 2, (11e) blocks the more general (11d) and suffixes -*er* to the stem. For the stems *big*, *bett* and *furth*, this results in the comparative forms *bigger*, *better* and *further*.

Rules (11a-e) account for the comparative realization of every lexeme except for BAD. In order to handle this suppletive item, Stump introduces a new device—the portmanteau rule in (11f). Portmanteau rules differ from other realizational rules in that they span more than one rule block. Rule (11f), for instance, operates over both Block 1 and Block 2. The effect is that, during rule competition, (11f) competes simultaneously with other applicable rules listed in either block. The relevant rules for this case are the defaults from each block—(11a) and (11d)—which apply to all adjectives. By Pāṇinian determinism, however, the more specific (11f) blocks both defaults, and the unsuffixed suppletive form *worse* is output.

Stump's (2001) analysis accounts for a number of important properties of the data. Different levels of productivity are modeled using realizational rules that apply to larger or smaller classes of words. The stem-selection rules for the comparative realization of the suppletives GOOD, BAD, and FAR, for example, amount to lexically-listed exceptions to the broader generalization that most comparatives are formed using an adjective's main stem. Likewise, rule (11d)—as the default in its block—can apply to any stem, while (11e) is limited to short adjectives. This captures relative differences in the productivity of the analytic and synthetic patterns.

Those who are unfamiliar with the realizational approach espoused by Stump (2001) might look at the rules in (11) and interpret them as incrementally building up the meanings associated with comparative forms. But it is not the case that the rules in Block 1 realize a stem-morpheme that is combined with an *-er* or *more* morpheme to form the comparative. On Stump's analysis, words have no internal constituent structure. The rules in (11) simply describe the ensemble of elements that make up comparative forms. For some words in some languages, pieces of this ensemble resemble canonical morphemes. But this outcome is not assumed as a grammatical principle. The only place in which forms (*better*, *farther*, *worse*, *bigger* and *more dangerous*) and meanings (lexeme-morphosyntactic feature set pairings) are associated is at the level of the paradigm cell.

This differs markedly from Embick's (2007) proposal. Both Embick's and Stump's analyses are uniform in the sense that responsibility for the creation of both synthetic and analytic forms is handled in the same linguistic module (the syntax for Embick, and morphology for Stump). The two accounts differ, however, in their commitment to the theoretical notion of the morpheme. In Embick's syntactic trees, meaningful stems are combined with the forms *-er* and *more* (which themselves carry a meaning: COMP/DEG). The meaning of a comparative is thus built up through the concatenation of smaller meaningful signs—i.e., morphemes. Nothing of the sort happens in Stump's analysis. As it turns out, this difference—whether or not morphemes are assumed to be primitives in morphology—has important implications for the modeling of speakers' intuitions about how morphologically complex words are. I turn to this topic in detail in §5.7.1.

As a member of the categorical family of comparative analyses (recall the taxonomy in Table 5.1), Stump's (2001) approach is—like Poser's (1992) and Embick's (2007)—committed to the Blocking Hypothesis for Competing Realizations. Like these others then, Stump is unable to account for optionality. This constitutes a significant hole in the empirical coverage his theory is able to provide.

5.7 Gradience in Grammar

While the analyses of Poser (1992), Embick (2007) and Stump (2001) all have very different linguistic pedigrees, they are identical in their neglect of the empirical evidence in favor optionality and, accordingly, of optionality's consequences for theory construction. Each author assumes the Blocking Hypothesis in (2), and rejects the Optionality Hypothesis in (3). This is symptomatic of a shared view of linguistic category structure—one in which category membership is viewed as an either-or proposition: any given adjective stem can be a member of either the synthetic class, or the analytic class, but not both. This view is an obvious oversimplification of the facts, and is probably related, in part, to the idea that comparative inflection can be accurately modeled by making reference to a single variable—stem length. If speakers' inflectional behavior is informed only by the number of syllables in a stem, then adjectives should partition neatly and deterministically into synthetic and analytic classes. On the other hand however, if speakers choose one comparative form over another based on the consideration of multiple, weighted factors—as the evidence suggests—then these factors should come into conflict, and lead to the same stem being probabilistically inflected in the synthetic class on some occasions, and the analytic class on others. I refer to this sort of non-deterministic categorization scheme as gradient.

Both the categorical and the gradient approaches to category structure have long histories. The categorical approach can be traced back to the philosophical writings of Aristotle, and later Frege (1960). Likewise, the idea that linguistic categorization is strict and deterministic has long found support in mainstream generative grammar—a paradigm that has "always been averse to the idea of fuzziness in grammar" (Aarts, 2004). Gradient approaches to categorization find their intellectual roots in Wittgenstein's philosophy of family resemblances (1968), Sapir's (1921) proclamation that "all grammars leak," and findings from prototype theory (Rosch, 1973) and cognitive linguistics (Lakoff, 1987; Langacker, 1987). More recently, gradient approaches have been embraced by a number of language researchers devoted to providing a more nuanced view of the language faculty than is often available under the assumption of categorical categorization (Baayen et al., 2003; Bochner, 1993; Boersma, 1997; Boersma & Hayes, 2001; Bresnan, 2006; Bresnan et al., 2007; Ernestus & Baayen, 2004; Hay, 2001; Hay & Baayen, 2005; Jaeger, 2006; Keller, 2000; Keller & Alexopoulou, 2001; Ramscar, 2002; Roland et al., 2006; Sorace & Keller, 2005).

For example, Keller and colleagues (Keller, 2000; Keller & Alexopoulou, 2001; Sorace & Keller, 2005) have repeatedly argued against the idea that sentence acceptability can be modeled using two discrete, non-overlapping categories labeled grammatical and ungrammatical. Linguists have traditionally assumed that the practice of collecting binary forced-choice judgments of acceptability is perfectly sufficient for theory construction. But the available evidence from a wide range of phenomena—Greek word order (Keller, 2000; Keller & Alexopoulou, 2001) and auxiliary selection in German and other Western European languages, among others (Keller, 2000; Sorace & Keller, 2005)—indicates that speaker judgments are much more nuanced, and appear to require modeling in a gradient system. In his (2000) thesis, for instance, Keller provides a taxonomy of constraints that govern the acceptability of extracted picture NPs (e.g., *Which friend has Thomas painted a picture of?*). These are divided into two classes: *hard* and *soft*. Keller shows that sentences that violate even a single hard constraint are judged as robustly unacceptable by native speakers. Hard constraints thus appear to be quite amenable to treatment in a categorical model of grammaticality. The violation of soft constraints, however, does not lead to robust ungrammaticality. Instead, there is a cumulative effect in which sentences that violate larger numbers of soft constraints are viewed as increasingly unacceptable. The finding that there are many subtle levels of acceptability between completely grammatical and completely ungrammatical cannot be accommodated in a system based on a categorical conception of grammaticality. A gradient model, on the other hand, has the flexibility to account for the way in which speakers react to violations of both soft and hard constraints.

The serious consideration of gradience has likewise had an impact in phonology, particularly in Optimality Theory (Kager, 1999; McCarthy & Prince, 1994; Prince & Smolensky, 1993). Boersma (1997) devised a stochastic version of Optimality Theory (OT) in which constraints are able to probabilistically overlap. This device allows the researcher to model categorical as well as gradient outcomes. For example, in Ilokano, an Austronesian language of the northern Philippines, underlying /ʔo/ sequences can be optionally realized as [ʔw], or [wʔ]: *pagbaʔwan* ~ *pagbawʔan*, ‘place where rats live’ (Hayes & Abad, 1989). Boersma and Hayes (2001) show that the appearance of one form over the other is regulated by the relative ranking of two constraints: a faithfulness constraint that favors [ʔw], and a markedness constraint that favors [wʔ]. In stochastic OT these constraints are allowed to commingle at roughly the same place on a continuous ranking scale. The effect is that *pagbaʔwan* is chosen as the optimal candidate on some evaluations, and *pagbawʔan* on others. Categorical outcomes are accounted for in the same model simply by positing that constraints are spaced further apart on the ranking scale. In this way, stochastic OT provides a unified system for handling a wide range of phenomena (Boersma & Hayes, 2001).

5.7.1 Gradient Complexity

Another example along these lines that has particular relevance for morphology comes from the study of morphological complexity. Speakers tend to intuitively feel that some words are more structurally complex than others. *Cats*, for example, is viewed as more complex than *cat*, *walked* as more complex than *walk*, and *laughable* as more complex than *laugh*. That different words seem to have different levels of structural complexity is not at issue. The question, rather, is whether complexity is a variable that can be binned deterministically into mutually exclusive categories such as *simple* (whose membership would include *cat*, *walk* and *laugh*) and *complex* (whose membership would include *cats*, *walked* and *laughable*)—or whether complexity is a gradient notion, subject to probabilistic constraints.

Carstairs-McCarthy (2005) notes that there are a number of interpretations concerning what counts as a sign (i.e., a form-meaning pairing) in morphology. Certainly words are signs. *Cat* and *cats*, for example, are undoubtedly forms that are paired with meanings. But theorists disagree about whether -*s* might be a sign as well. According to a *morpheme-based* view of morphology, the independent word *cat* is formally and semantically unified with the sign -*s* to form *cats*. Both *cat* and -*s* count as morphological primitives known as *morphemes* that form the building blocks for complex words such as *cats*. On this view, *cats* does not have independent status in the grammar. Its psychological standing as a word of English is the epiphenomenal result of the combination of *cat* with -*s*. The opposing view claims that *cat* and *cats* have independent grammatical status, but not -*s*. Such approaches are referred to as *word-based*, and hold that, over the course of using singular-plural word pairs such as *cat*-*cats* and *dog*-*dogs*, speakers may epiphenomenally come to associate -*s* with PLURAL. It is not the case, however, that the grammar grants -*s* independent status as a sign (as in morpheme-based theory).

These different approaches to morphology make quite different predictions with respect to word complexity. If words are built up out of morphemes, as morpheme-based theory holds, then complexity must be categorical. A word will be simple if it is composed of a single morpheme (e.g., *cat*), and complex if it is composed of multiple morphemes (e.g., *cats*). No middle ground is predicted—it is nonsensical under the morpheme-based view to say that *cats* is *somewhat* complex, since this would entail that it was unclear whether *cats* consists of one or two morphemes.

The alternative prediction made by word-based theory is, of course, that structural complexity is graded (Hay, 2001; Hay & Baayen, 2001, 2005). This hypothesis finds support in a variety of studies (Milin et al., 2007; Moscoso del Prado Martín et al., 2004), but most notably in an experiment by Hay (2001) in which subjects were presented with word pairs, and were asked to choose the member of each pair that was most easily decomposable into constituent parts. This is akin to asking subjects to identify the word in each pair that they felt was more complex. All of the items used by Hay were unambiguously complex on a morpheme-based view of morphology. *Adornment* and *alignment*, for example, are both analyzed as consisting of the nominalizing morpheme -*ment* suffixed to the verbal stems *adorn* and *align*. The items in each of Hay's pairs differed, however, according to the relative frequency of each item to its associated stem. A*dornment*, for example, has a CELEX frequency of 41, while *adorn* has a frequency of 75 (Baayen, Piepenbrock, & Gulikens, 1995). This amounts to an *adornment* to *adorn* frequency ratio less than one. The frequency ratio of *alignment* to *align*, however, is greater than one: *alignment*'s frequency is 57, and *align*'s is 44. Hay found that this frequency manipulation accounted for a significant amount of variability in her subject's responses. Given a word pair in which one member had a frequency ratio less than one (e.g., *adornment*) while the other had a frequency ratio greater than one (*alignment*), subjects reliably found the item with the lower frequency ratio to be the more complex member of the pair (e.g., *adornment* was viewed as more complex than *alignment*).

This result is important because it illustrates the pivotal role that the comparison of whole words can play in explaining linguistic behavior. Specifically, the perception of complexity is tied to the frequencies associated with the lexical entries of paradigmatically-related words. Morpheme-based approaches are not designed to account for this pattern of results because complex words do not have lexical entries that can be associated with frequency counts. That is, on a morpheme-based account, only *adorn*, *align*, and -*ment* are lexically listed. And since frequency is associated with lexical items, only these morphemes can have frequencies. *Adornment* and *alignment*, on the other hand, are the output of a morphological process that adds -*ment* to listed verbs, so neither word has any independent lexical status. What this means is that morpheme-based theory has no way to derive real complexity intuitions like the ones Hay (2001) reports, because it cannot represent *adornment* and *alignment* lexically, and so lacks a means for storing their frequencies for subsequent comparison with those of *adorn* and *align*.

Hay's (2001) findings agree with those noted above from Keller and colleagues (Keller, 2000; Keller & Alexopoulou, 2001; Sorace & Keller, 2005), and the stochastic OT literature (Boersma, 1997; Boersma & Hayes, 2001; Boersma & Levelt, 2000). A number of results thus appear to converge on the same conclusion: in order to accurately account for data that show categorical as well as gradient effects, it is necessary to reject a purely categorical view of category structure, and replace it with a more flexible, gradient approach. With respect to comparative inflection, this way of thinking suggests that the Blocking Hypothesis (and all analyses that adopt it, e.g. Poser 1992, Embick 2007 and Stump 2001) should be rejected in favor of the Optionality Hypothesis.

Moreover, the gradient approach to morphological complexity espoused by Hay and colleagues (Hay, 2001; Hay & Baayen, 2001, 2005), interfaces naturally with word-based theories of morphology. The fact that complexity intuitions can be derived through the comparison of the frequencies associated with whole-word forms suggests that paradigmatic contrast may have a crucial role to play as an explanatory construct in morphological theory. Evidence in favor of the psychological reality of paradigms has been mounting in recent experimental studies (Baayen, Feldman, & Schreuder, 2006; Janssen & Penke, 2002; Kostić, Marković, & Baucal, 2003; Milin, Đurđević, & Moscoso del Prado Martín, 2007; Moscoso del Prado Martín, Baayen, & Kostić, 2004). Most of these works find that metrics that incorporate some notion of the paradigm predict a significant amount of variance in processing times (as in, for example, a lexical decision task). The recognition of paradigms as theoretical constructs is therefore warranted because they assist in explaining outcomes on a variety of experimental measures—lexical decision times and complexity judgments (Hay, 2001)—and are amenable to gradient treatments of grammar (Hay & Baayen, 2005) in a way that the morpheme is not.

5.8 Word and Paradigm Morphology

Word-based morphology—also known as Word and Paradigm (WP) morphology—has a long history dating back to the Greek and Latin grammarians (Matthews, 1991). In comparison, morpheme-based morphology is a relatively recent invention that is canonically associated with research coming out of the post-Bloomfieldian American structuralist tradition (Bloomfield, 1949). Arguments in favor of WP morphology are thus arguments in favor of a return to a more well-established method of morphological analysis (Ackerman & Malouf, 2007; Ackerman & Stump, 2004; J. P. Blevins, 2005, 2006; Chomsky, 1965; Matthews, 1991; Spencer, 2004; Stewart & Stump, 2007; Stump, 2001, 2002).

As an introduction to WP morphology, consider the following table of verbforms from Spanish. These are quoted from Gurevich (2006), and originate in Matthews (1991).

Table 5.2: Paradigmatic contrast marking in Spanish

	INDICATIVE	SUBJUNCTIVE	INDICATIVE	SUBJUNCTIVE
1SG	compro	compre	como	coma
2SG	compr**a**s	compr**e**s	com**e**s	com**a**s
3SG	compr**a**	compr**e**	com**e**	com**a**
1PL	compr**a**mos	compr**e**mos	com**e**mos	com**a**mos
2PL	compr**á**is	compr**é**is	com**é**is	com**á**is
3PL	compr**a**n	compr**e**n	com**e**n	com**a**n

Spanish verbs in the *-ar* conjugation class (e.g., *comprar*, 'to buy') distinguish between the indicative and subjunctive moods through the use of a contrast in the suffix vowel (in bold, above). Forms in the indicative tend to be marked with an *-a*, whereas forms in the subjunctive tend to be marked with an *-e*. A morpheme-based analysis might mistakenly seize on this fact and declare that *-a* and *-e* are morphemes in Spanish that carry the meanings INDICATIVE and SUBJUNCTIVE, respectively. This error would hopefully be short-lived though, since consideration of the paradigm of any verb in the -*er* conjugation class (e.g., *comer*, 'to eat') would indicate that -*e* and -*a* cannot possibly be associated with those meanings: for -*er* verbs, -*e* occurs in indicative forms, and -*a* in subjective forms.

This lesson illustrates one of the central tenets of WP theories of morphology: morphemic analysis often fails. While this is clearly the case for Spanish, it is also true in highly agglutinative languages such as Turkish and Hungarian, which purportedly—in the view of morpheme theorists—show strong support in favor of the morpheme (Spencer, 2004). In practice, the only level at which stable form-meaning associations can be assessed is that of the *word*. This observation has been demonstrated time and again through the careful cross-linguistic study of regularly-occurring deviations from the assumption in morpheme-based theories that biunique form-meaning mappings are universal, and that their locus is the morpheme. Examples of *extended exponence* (one meaning mapping to multiple pieces of a word) have been illustrated in Latin (Matthews, 1981) and Breton (Stump, 2001), among others. Likewise, examples of *fusion* (multiple meanings mapping to one exponent) have been widely documented in Indo-European nominal inflection (Spencer, 2004). Still other deviations from biuniqueness include *zero-morphs*, and *meaningless morphs*. Zero morphs occur in morpheme theory whenever a word is interpreted as carrying a morphosyntactic property without having any overt marking of that property. A common example is the English singular noun. On a morpheme-based theory, plurals like *cars* get their number marking through -*s* suffixation. The stem *car*, on the other hand, is inherently numberless and can only be marked for singular by adding a zero-suffix, as in *car-Ø*. Meaningless morphs are exemplified by the Latinate formations *refer*, *confer*, *defer* and *receive*, *conceive*, and *deceive*, as discussed by Aronoff (1976). These forms seem to indicate a segmentation into the prefixes *re-*, *con-* and *de-*, and the roots *-fer* and *-ceive*. None of these formatives have discernable meanings though.

As argued in Chomsky (1965), slavish adherence to the morpheme as a theoretical construct can have undesirable consequences. Embick (2007; see discussion in §5.5), for instance, combines a syntactocentric DM approach with a morphemic analysis. Deg and Adj are treated as meaningful pieces (i.e., morphemes) that are moved around in a syntactic tree alongside phonetically unrealized—and arguably unmotivated—morphemes like the κ in Embick's κP.

The kinds of byzantine maneuvering that are necessary under morpheme-based theory to maintain the assumption that each piece of a word has a single, discernable meaning are not required when one adopts the position that words are the true locus of meaning in morphology. According to this view, meaning is ascribed to words as a result of their juxtaposition with other items in paradigms. Returning to the example in Table 5.2, *compre* is interpreted as bearing the morphosyntactic property SUBJUNCTIVE not because it carries a morpheme that bears that meaning, but because it is a member of a paradigm (i.e., the paradigm for *comprar*), which is itself a member of a large conjugational class. If a third-person singular form occurs with an *-e* in an *-ar* paradigm, then it has a subjunctive interpretation. The same *-e* exponent in an *-er* paradigm, however, means that the form is, by hypothesis, indicative. Native speakers of Spanish know the correct interpretation for a host of different forms because they have knowledge of each word's place within a wider system of meaning.

5.9 A Pattern-Based Gradient-Uniform Comparative Model

In this section, I describe a gradient, pattern-based model of comparative inflection by Bochner (1993). This model is consistent with WP assumptions concerning the centrality of paradigmatic relations among whole wordforms to morphological explanation. Moreover, it is compatible with the Optionality Hypothesis for Competing Realizations, introduced in (3), and repeated below.

(12) Optionality Hypothesis for Competing Realizations

$$P(e_n) = [0, 1]$$

According to the Optionality Hypothesis, competing realizations are not arrayed categorically such that the expression of one precludes the expression of another. Analyses that follow the categorical approach—i.e., that adhere to the Blocking Hypothesis—have been argued in §5.4-5.6 to be inadequate. Bochner (1993) instead provides a unique perspective on comparative optionality—one that allows single adjective stems to be inflected according to both the synthetic and analytic patterns. This is accomplished by positing a rich set of morphological patterns that speakers are able to work from. Like Stump (2001), Bochner's analysis does not address analytic inflection directly. There is nothing, however, in the framework that Bochner assumes that precludes dealing with analytic and synthetic expressions in the same unified morphological system. The version of Bochner that I describe below is thus slightly modified from the one that he himself presents.

Bochner's (1993) analysis is summarized by the rules given in (13). Each rule describes a lexicalized pattern that relates two paradigmatically-related words. Rules in the left-hand column describe patterns of synthetic inflection. The single rule in the right-hand column describes analytic inflection.

(13) Summary of Bochner (1993)

SYNTHETIC

a. $$\begin{bmatrix} /\mathrm{X}/ \\ \mathrm{Adj} \\ \mathrm{Z} \end{bmatrix} \leftrightarrow \begin{bmatrix} /\mathrm{X}\textit{er}/ \\ \mathrm{Adj} \\ \mathrm{MORE\ Z} \end{bmatrix}$$

c. $$\begin{bmatrix} /\sigma/ \\ \mathrm{Adj} \\ \mathrm{Z} \end{bmatrix} \leftrightarrow \begin{bmatrix} /\sigma\textit{er}/ \\ \mathrm{Adj} \\ \mathrm{MORE\ Z} \end{bmatrix}$$

d. $$\begin{bmatrix} /\mathrm{X}\textit{y}/ \\ \mathrm{Adj} \\ \mathrm{Z} \end{bmatrix} \leftrightarrow \begin{bmatrix} /\mathrm{X}\textit{ier}/ \\ \mathrm{Adj} \\ \mathrm{MORE\ Z} \end{bmatrix}$$

ANALYTIC

b. $$\begin{bmatrix} /\mathrm{X}/ \\ \mathrm{Adj} \\ \mathrm{Z} \end{bmatrix} \leftrightarrow \begin{bmatrix} /\textit{more}\ \mathrm{X}/ \\ \mathrm{Adj} \\ \mathrm{MORE\ Z} \end{bmatrix}$$

Materials enclosed within brackets represent schematic lexical entries. The first element in each entry is a phonological form (written between forward slashes), the second element gives the entry's lexical category, and the third element (in small caps) briefly represents its meaning. Each rule describes an association between an uninflected adjective stem and its corresponding comparative. The rule in (13a), for instance, relates a lexical item with form X, part of speech Adj, and meaning Z to another lexical item with form X*er*, part of speech Adj, and the meaning MORE Z. The two-way arrow in each rule can be thought of as symbolizing a mutual licensing arrangement between lexical entries. That is, if a lexicon contains a word that matches the specification on the left side of rule, then the grammar licenses its relative on the right side. Likewise, if the lexicon contains a word that matches the specification on the right side, then the grammar licenses the related word on the left. The rules in (13a) and (13b) can be thought of as general specifications in the grammar of the synthetic and analytic patterns, respectively. Given any adjective stem, they license a comparative of the form X*er*, or *more* X. The rules in (13c) and (13d) are more constrained versions of the general synthetic rule. (13c) relates any monosyllabic adjective stem to a comparative consisting of a monosyllabic stem followed by *-er* (e.g., *strong* ↔ *stronger*). Likewise, (13d) relates any adjective stem ending in *–y* (regardless of length) to a comparative form ending in *-ier* (e.g., *persnickety* ↔ *persnicketier*).

There are three properties of the analysis in (13) that are noteworthy. First, in no sense are the comparatives that are allowed in this system formed by concatenating meaningful formal chunks, as in a morpheme-based analysis. Instead, meanings are assigned at the word-level (or the multi-word level, in the case of the analytic). Second, if a stem fits the licensing restrictions in the left side of any rule, then that rule is able to apply to form a comparative. Since rules (13a) and (13b) can apply to any adjective stem, this entails that optionality is built into the system (it additionally predicts rampant overgeneralization, but this issue is addressed below in §5.9.1). Third, oddly, there appears to be a certain amount of redundancy with respect to synthetic realization. Specifically, why are (13c) and (13d) needed when the data that they account for is covered by the more general rule of synthetic inflection in (13a)?

This question points to a fundamental difference between Bochner's (1993) analysis and the others that have been reviewed above (i.e., Embick, 2007; Poser, 1992; Stump, 2001). All analysts (including Bochner) attempt to provide a characterization of comparative inflection that is maximally simple. But Bochner differs from the others in terms of how simplicity is defined. Note in Poser (1992), for example, that the analysis makes reference to only two patterns: one that suffixes *-er* to short adjectives, and one that combines *more* with all other adjective stems. This account is maximally simple in its *brevity*. That is, Poser does not refer to other patterns in the data—e.g., the fact that *-y* endings tend to be amenable to suffixation (Graziano-King, 2003)—but instead gives an analysis that condenses comparative inflection down to its bare essentials. Bochner (1993) identifies this notion of simplicity as adhering to what he calls the *symbol-counting evaluation metric* (Halle, 1962). According to the symbol-counting metric, maximally simple grammars are those that posit the fewest rules.

Bochner (1993), however, argues that the symbol-counting metric is inadequate, and that additional rules only contribute to the complexity of the grammar to the extent that they add *independent information* to it. Since (13c) and (13d) are not independent of (13a)—they are only more highly-specified variants of it—they do not add much independent information, and thus keep the complexity of the analysis to a minimum. Further, more highly-specified rules like (13c) and (13d) are justified because they provide a tighter fit to the data than (13a) does. A general rule of suffixation—like (13a)—predicts overgeneralization errors of the kind discussed in Chapter 4, e.g., **dangerouser*. That errors of this sort do not occur in adult speech is evidence that more constrained generalizations—such as (13c) and (13d)—are in use.

The notion that simplicity is related to the amount of independent information in a grammar finds support in a recent computational model of the English past tense. Albright and Hayes (2003) devised an inductive learning algorithm (Albright & Hayes, 2002) that discovered rules governing past tense formation. These rules were assigned *reliability* scores (Bochner refers to the same concept as *regularity*). Roughly, reliability is the number verbs in the dataset that a rule actually applies to, divided by the number of verbs to which it could apply. Albright and Hayes' algorithm discovered dozens of rules, each with different levels of reliability. Abbreviated representations for two of these are given in (14), along with their reliability scores.

(14) Reliability in Albright and Hayes (2003)

	RULE	RELIABILITY
a.	$X \rightarrow X\textit{ed}_{[\text{+PAST}]}$	4034 / 4253 = 0.95
b.	$X_{[\text{-SONORANT, +CONTINUANT, -VOICE}]} \rightarrow X[t]_{[\text{+past}]}$	352 / 352 = 1.00

(14a) is a general suffixation rule for regular past tense inflection: given any input verb stem, the rule forms the past tense by adding *-ed* (e.g., *walk* → *walked*). According to Albright and Hayes' dataset, this rule has high reliability: it applies to 95% of verb stems that it could possibly apply to. The rule in (14b), however, is even more reliable. It suffixes the [t] allophone to verb stems that end in a voiceless fricative (e.g., *laugh* → *laughed*), and applies to 100% of stems ending in voiceless fricatives. Since (14a) is more broadly applicable than (14b), adherents of the symbol-counting evaluation metric (e.g., Pinker & Prince, 1988) would prune (14b) from the grammar, and retain only the maximally general (14a). Albright and Hayes (2003), however, argue against this course of action. They demonstrate experimentally that speakers have stronger, more accurate intuitions about forms that are derived using relatively more reliable rules, versus those that are derived use less reliable rules. This counts as evidence that even narrow rules like (14b) have psychological validity, and should thus be included in models of speaker competence.

The same line of reasoning applies to the comparative-formation rules in (13c) and (13d). While Bochner (1993) does not explicitly calculate reliability scores for his three rules of synthetic formation, he argues that (13c) and (13d) necessarily fit the data better (i.e., are more reliable) than the overly broad (13a). They should therefore be included in the analysis. Doing so has the effect of reconceptualizing grammar. Speakers' knowledge of comparative inflection is conceived of as consisting of multiple generalizations, sometimes related, that are captured in the lexicalized patterns in (13). This knowledge cannot be adequately represented using the maximally brief formulations that are preferred under the symbol-counting evaluation metric.

5.9.1 A Probabilistic Revision of Bochner (1993)

While I regard Bochner's (1993) approach to comparative inflection as a significant improvement over the categorical analyses described in §5.4-5.6, it is not entirely unproblematic. In this section I outline deficiencies in Bochner's proposal, and suggest ways in which they could be remedied. The outcome will be a gradient, word-based analysis of comparative inflection that probabilistically selects between the synthetic and analytic patterns based on the consideration of multiple, weighted factors. This revised analysis is compatible with results from the statistical modeling literature in that linguistic behavior is conceptualized as having multiple causes (Bresnan, 2006; Bresnan et al., 2007; Jaeger, 2006; Roland et al., 2006). It differs, however, in that it embeds statistical tendencies within a grammatical theory that treats linguistic patterns as lexicalized morphological and syntactic constructions (Ackerman & Malouf, 2007; Goldberg, 1995, 2006; Kay & Fillmore, 1999).

To begin, the set of rules proposed in (13) is insufficient to account for the broad array of formal determinants that have been discussed in this thesis. Bochner's (1993) analysis really only makes reference to phonological properties that are known to predict comparative form: syllabicity and the presence of *-y* stem endings. On the one hand, this is somewhat odd given that Bochner spends a significant amount of time arguing that conceptualizing complexity as independent information rather than brevity allows him to posit a large number of rules. If the addition of non-independent rules such as (13c) and (13d) does not increase the complexity of the grammatical description, then many more rules should be allowed. The truth, though, is that only phonological variables are mentioned because many of the predictors of comparative form that we now know about had not been identified at the time that Bochner was writing. Their absence from his analysis is therefore quite understandable. That we now, however, have a firmer understanding of the effects on comparative inflection of complexity, frequency, gradability and syntactic class, motivates adding to the four rules in (13).

A more serious shortcoming of Bochner's (1993) formulation, however, is that there is no mechanism to determine, on a given speech event, which rule applies to an input among all those that are eligible to apply. Bochner's analysis amounts to a statement of generalizations that are *possible* in the comparative domain, without saying anything about which might be more *probable*. The rules in (13) can be thought of as an inventory of lexicalized patterns that exist in the adult grammar. But there is nothing that determines which patterns are used where. As a result, Bochner's model potentially over-generates forms. The two general rules of synthetic and analytic inflection—(13a) and (13b) predict that any adjective stem can be inflected using either pattern. This kind of rampant optionality is obviously not a part of native-speaker behavior, however.

The problem of overgeneralization can be straightforwardly solved by adopting a technique used by Albright and Hayes (2003) to constrain the generative capability of their model—namely, annotating the set of stored patterns for reliability. Albright and Hayes' experiments suggest that reliability is a good predictor of production probability. In production, speakers are more likely to rely on patterns that they have found to be highly reliable, versus those that are less reliable. This provides a probabilistic method for choosing among competing patterns during a production event. As an example, consider the inflection of the monosyllabic stem *old*. Three rules potentially apply to *old*: (13a) and (13c) predict *older* as an output, and (13b) predicts **more old*. Table 5.3 gives estimated reliability scores for the three relevant rules.

Table 5.3: Estimated reliability scores

RULE	RELIABILITY
13a	20%
13b	80%
13c	> 80%

I use type frequencies reported in Boyd (2006) to estimate the reliability scores for (13a) and (13b). Since corpus counts indicate a one-to-four type frequency ratio of synthetically inflected adjectives to analytically inflected adjectives, the general rules of synthetic and analytic inflection should have reliability scores around 20% and 80%, respectively. I assume, based on the strong correlation between monosyllabicity and synthetic inflection, that the reliability of (13c) will be somewhat higher than 80%. These numbers ensure that *old* will be inflected on the majority of occasions according to the pattern in (13c), as *older*.[13] (13c) wins the competition because, among the rules that are able to apply to *old*, it has the highest reliability.

The final problem with Bochner's (1993) analysis is that it often does not posit rules that are sufficiently constrained. As things currently stand, assuming the reliability scores in Table 5.3, monosyllabic stems would always be favored to inflect synthetically, under rule (13c), because (13c) is more reliable than (13b), the analytic rule. This is because (13c) is too broadly applicable. Experimental data suggest that a cluster of features combine to produce synthetic inflection in monosyllabics: Graziano-King (2003), for example, shows that only monosyllabic stems that are high frequency and gradable tend to inflect with *-er*. Revising (13c) to reflect this fact would probably increase its reliability. It would also, however, constrain (13c) sufficiently so that it would not be able to apply to exactly those monosyllabic stems that tend towards analytic inflection: low-frequency non-gradables such as *right* (**righter*), and *just* (**juster*). The model should thus be able to approximate speaker behavior more accurately as the rules are adjusted to give a tighter fit to the data.

In summary, a revised version of Bochner's (1993) model meets all of the criteria that have been suggested throughout the course of this work as being desirable properties of a grammatical analysis of comparative inflection. Sensitivity to multiple determinants of comparative form can be built into the model, optionality is allowed, and unsupported

[13] I currently have no means of modeling exactly how often a rule with a higher reliability score will be chosen over one with a lower score. I assume though, borrowing from stochastic OT (Boersma, 1997; Boersma & Hayes, 2001; Boersma & Levelt, 2000), that this will be a function of the difference between the two scores, with larger differences being associated with more categorical outcomes, and smaller differences with optionality.

claims concerning the syntactic nature of analytics (Poser, 1992), and analytics and synthetics (Embick, 2007) are rejected. The analysis is additionally consistent with construction-based theories of language, in which linguistic competence derives from the facile use of many stored patterns (Ackerman & Malouf, 2007; Goldberg, 1995, 2006; Kay & Fillmore, 1999). Here, I focus on morphological word-based patterns, but there is nothing in the broader framework that prevents syntactic patterns from being dealt with in the same way. Finally, I suggest that recent statistical models of linguistic behavior (Bresnan, 2006; Bresnan et al., 2007; Jaeger, 2006; Roland et al., 2006) are accommodated straightforwardly in a constructional, pattern-based approach. Models of this sort point to a host of variables that speakers consider when formulating their utterances. To the extent that these variables are useful predictors of comparative form, they can be formulated as rules, annotated for reliability, and included in a Bochnerian pattern-based model of inflection.

5.10 Concluding Remarks

The goal of the present work as been to use the methods provide by experimental psychology and grammatical analysis to inaugurate the psycholinguistic study of morphological optionality. The study's findings include results from comprehension, production, and acquisition. In the comprehension work described in Chapter 2, I demonstrate that among adjectives that participate in optional comparative inflection, listeners roundly prefer the analytic variant. The available evidence suggests that this outcome follows from an analytic parsing advantage that makes phrase structure available eariler, and avoids a temporary parsing ambiguity associated with the synthetic.

In a production experiment conducted jointly with L. Robert Slevc and described in Chapter 3, we find that speakers tend to increase their rate of analytic use under conditions of syntactic and semantic complexity. There is no evidence that this behavior results from an attempt to ease the processing burden on addressees. Instead, increased use of the analytic is consistent with findings that indicate that speaker choice more often reflects constraints on speakers' own production processes. Chapter 4 details work conducted on the acquisition of comparative inflection. There, I demonstrate that children overgeneralize both the synthetic and analytic inflectional patterns, and show that previous researchers may have failed to document this behavior because they did not adequately control for response perseveration. In Chapter 5, I introduce the Blocking and Optionality Hypotheses for Competing Realizations, and argue that a wide range of linguistic analyses of the comparative are inadequate because they assume the Blocking Hypothesis, and are consequently unable to model systematic variability in comparative form. Adopting the Optionality Hypothesis, on the other hand, gives researchers the flexibility to account for categorical and gradient effects within the same integrated framework. A pattern-based analysis that is consistent with Word and Paradigm assumptions as well as the Optionality Hypothesis is suggested in §5.9.

Together, these results indicate that morphological optionality is subject to many of the same constraints as syntactic optionality. While listeners may prefer the analytic variant, speakers choose to rely on whatever pattern of inflection suits their current needs. In many cases, the synthetic may be preferred because it is supported by significantly higher token frequencies, and should therefore gain an advantage in terms of ease of activation. Speakers, however, are also capable of flexibly choosing between the synthetic and the analytic patterns, and appear to increasingly rely on the analytic as the complexity of the utterance they are attempting to encode increases. Moreover, morphological patterns appear to follow the same kind of developmental trajectory as syntactic patterns. The evidence suggests that the comparative system is constructed based on abstraction over experience with specific items. The resulting representations are initially much more general than those of the adult grammar. With time, however, children's behavior becomes increasingly adult-like. That morphological and syntactic patterns show processing similarities at the adult stage, and developmental similarities over the course of acquisition suggests the viability of a construction-based approach that treats words and phrases as different parts of the same system.

Appendix A: Critical Sentences in Experiments 1 and 2

1. Although having to pay the speeding ticket made Elaine mad, she was angrier to find that it had affected her insurance premiums.
2. Although having to pay the speeding ticket made Elaine mad, she was more angry to find that it had affected her insurance premiums.
3. Children with Howson’s syndrome are apter to react inappropriately to antibiotics.
4. Children with Howson’s syndrome are more apt to react inappropriately to antibiotics.
5. While bicycling without a helmet is a bit foolish, it’s crazier to ride a motorcycle without one.
6. While bicycling without a helmet is a bit foolish, it’s more crazy to ride a motorcycle without one.
7. The veterinarian told Mary that it would be crueler to allow her cat to continue suffering than to put it to sleep.
8. The veterinarian told Mary that it would be more cruel to allow her cat to continue suffering than to put it to sleep.
9. Some news items are fitter to print than others.
10. Some news items are more fit to print than others.
11. The tax code changes leave Americans freer to manage their money than ever before.
12. The tax code changes leave Americans more free to manage their money than ever before.
13. After the vacation she was feeling relaxed, refreshed, and hungrier to resume work than she had been for some time.
14. After the vacation she was feeling relaxed, refreshed, and more hungry to resume work than she had been for some time.
15. Some brokers are keener to defend against a loss than to secure a gain.
16. Some brokers are more keen to defend against a loss than to secure a gain.
17. The 2004 Ford GT, with it’s newly-redesigned suspension system, is livelier to drive than ever before.
18. The 2004 Ford GT, with it’s newly-redesigned suspension system, is more lively to drive than ever before.
19. Local legend has it that a four-leafed clover is luckier to find on St. Patrick’s Day than on any other day of the year.
20. Local legend has it that a four-leafed clover is more lucky to find on St. Patrick’s Day than on any other day of the year.
21. Kenny G is mellower to listen to than Yanni.
22. Kenny G is more mellow to listen to than Yanni.
23. Highway 95 is pleasanter to drive during the summer months.
24. Highway 95 is more pleasant to drive during the summer months.
25. Armstrong was prouder to prevail against cancer than to win against all of his cycling opponents.
26. Armstrong was more proud to prevail against cancer than to win against all of his cycling opponents.

27. Young people today are much readier to express themselves in public.
28. Young people today are much more ready to express themselves in public.
29. The road from San Diego to San Bernardino is riskier to drive if it has recently rained.
30. The road from San Diego to San Bernardino is more risky to drive if it has recently rained.
31. Football fans in Los Angeles were sorrier to see the Raiders leave for Oakland than to see the Rams leave for St. Louis.
32. Football fans in Los Angeles were more sorry to see the Raiders leave for Oakland than to see the Rams leave for St. Louis.
33. As a result of the economic downturn, the fourth-quarter earnings numbers were starker to review than anyone had anticipated.
34. As a result of the economic downturn, the fourth-quarter earnings numbers were more stark to review than anyone had anticipated.
35. Nothing is surer to rally students against the administration than a hike in fees.
36. Nothing is more sure to rally students against the administration than a hike in fees.

Appendix B: Critical Sentences in Experiment 3

1. The cop was angrier than the sailor.
2. The cop was angrier to hear that the band was breaking up than the sailor.
3. The cop was more angry than the sailor.
4. The cop was more angry to hear that the band was breaking up than the sailor.
5. Given the current political climate, the Ambassador’s comments were apter than the Senator’s.
6. Given the current political climate, the Ambassador’s comments were apter to incite violence than the Senator’s.
7. Given the current political climate, the Ambassador’s comments were more apt than the Senator’s.
8. Given the current political climate, the Ambassador’s comments were more apt to incite violence than the Senator’s.
9. Hummers are costlier than Volvos.
10. Hummers are costlier to own and operate than Volvos.
11. Hummers are more costly than Volvos.
12. Hummers are more costly to own and operate than Volvos.
13. Mike was crazier than Eddie.
14. Mike was crazier to turn down the job than Eddie.
15. Mike was more crazy than Eddie.
16. Mike was more crazy to turn down the job than Eddie.
17. Russian winters are crueler than Canadian winters.
18. Russian winters are crueler to experience than Canadian winters.
19. Russian winters are more cruel than Canadian winters.
20. Russian winters are more cruel to experience than Canadian winters.
21. Kung fu is deadlier than karate.
22. Kung fu is deadlier to behold than karate.
23. Kung fu is more deadly than karate.
24. Kung fu is more deadly to behold than karate.
25. Lance Armstrong is fitter than Hugh Hefner.
26. Lance Armstrong is fitter to run a marathon than Hugh Hefner.
27. Lance Armstrong is more fit than Hugh Hefner.
28. Lance Armstrong more fit to run a marathon than Hugh Hefner.
29. The glass vase looks frailer than the metal one.
30. The glass vase looks frailer to touch than the metal one.
31. The glass vase looks more frail than the metal one.
32. The glass vase looks more frail to touch than the metal one.
33. The unchained tiger is freer than the chained tiger.
34. The unchained tiger is freer to roam as he pleases than the chained tiger.
35. The unchained tiger is more free than the chained tiger.
36. The unchained tiger is more free to roam as he pleases than the chained tiger.
37. The boy is hungrier than the man.
38. The boy is hungrier to eat Cheerios than the man.
39. The boy is more hungry than the man.

40. The boy is more hungry to eat Cheerios than the man.
41. Some stockbrokers are keener than others.
42. Some stockbrokers are keener to defend against a loss than others.
43. Some stockbrokers are more keen than others.
44. Some stockbrokers are more keen to defend against a loss than others.
45. Horses are livelier than mules.
46. Horses are livelier to ride than mules.
47. Horses are more lively than mules.
48. Horses are more lively to ride than mules.
49. Four-leafed clovers are luckier than pennies.
50. Four-leafed clovers are luckier to find than pennies.
51. Four-leafed clovers are more lucky than pennies.
52. Four-leafed clovers are more lucky to find than pennies.
53. Kenny G is mellower than Yanni.
54. Kenny G is mellower to listen to than Yanni.
55. Kenny G is more mellow than Yanni.
56. Kenny G is more mellow to listen to than Yanni.
57. Highway 55 is pleasanter than Highway 93.
58. Highway 55 is pleasanter to drive than Highway 93.
59. Highway 55 is more pleasant than Highway 93.
60. Highway 55 is more pleasant to drive than Highway 93.
61. Marines are prouder than sailors.
62. Marines are prouder to fight for their country than sailors.
63. Marines are more proud than sailors.
64. Marines are more proud to fight for their country than sailors.
65. Well-prepared students are readier than slackers.
66. Well-prepared students are readier to express themselves than slackers.
67. Well-prepared students are more ready than slackers.
68. Well-prepared students are more ready to express themselves than slackers.
69. In circus acts, chainsaws are usually riskier than knives.
70. In circus acts, chainsaws are usually riskier to juggle than knives.
71. In circus acts, chainsaws are usually more risky than knives.
72. In circus acts, chainsaws are usually more risky to juggle than knives.
73. The water in Kenya was saltier than the food.
74. The water in Kenya was saltier to taste than the food.
75. The water in Kenya was more salty than the food.
76. The water in Kenya was more salty to taste than the food.
77. Teddy was sorrier than Brian.
78. Teddy was sorrier to have broken the window than Brian.
79. Teddy was more sorry than Brian.
80. Teddy was more sorry to have broken the window than Brian.
81. Goya’s work is generally starker than Picasso’s.
82. Goya’s work is generally starker to behold than Picasso’s.
83. Goya’s work is generally more stark than Picasso’s.
84. Goya’s work is generally more stark to behold than Picasso’s.

85. When it comes to trivia, Susan is surer than George.
86. When it comes to trivia, Susan is surer to know the answers than George.
87. When it comes to trivia, Susan is more sure than George.
88. When it comes to trivia, Susan is more sure to know the answers than George.
89. Smith was unhappier than Jenkins.
90. Smith was unhappier to be fired than Jenkins.
91. Smith was more unhappy than Jenkins.
92. Smith was more unhappy to be fired than Jenkins.
93. The Patagonia jacket is warmer than the North Face jacket.
94. The Patagonia jacket is warmer to wear than the North Face jacket.
95. The Patagonia jacket is more warm than the North Face jacket.
96. The Patagonia jacket is more warm to wear than the North Face jacket.

Appendix C: Experiment 4 Prompts

1. - The Village People were upset.
   - 16% of music fans thought the sailor was angry to hear that the band was breaking up.
   - 76% of music fans thought the cop was angry to hear that the band was breaking up.
2. - The Village People were upset.
   - 16% of music fans thought the sailor was angry.
   - 76% of music fans thought the cop was angry.
3. - Joe and Brian were doing a project.
   - 24% of professors believed Joe’s ideas were complete.
   - 56% of professors thought that Brian's ideas were complete.
4. - Some pieces may have been missing from some puzzles in a toy store.
   - 24% of employees thought that the first puzzle was complete.
   - 56% of employees thought that the second puzzle was complete.
5. - Mechanics were asked about new vehicles.
   - 30% of mechanics said that Volvos are costly to own and operate.
   - 90% of mechanics said that Hummers are costly to own and operate.
6. - Mechanics were asked about new vehicles.
   - 30% of mechanics said that Volvos are costly.
   - 90% of mechanics said that Hummers are costly.
7. - Ambassadors were asked about international relations.
   - 56% of ambassadors said that Egypt and Israel are cozy.
   - 67% of ambassadors said that the U.S. and Israel are cozy.
8. - School kids were asked about winter clothes.
   - 56% of kids said that gloves are cozy.
   - 67% of kids said that mittens are cozy.
9. - Both Mike and Eddie turned town a prestigious job.
   - 23% of their friends thought that Eddie was crazy to turn down the job.
   - 89% of their friends thought that Mike was crazy to turn down the job.
10. - Both Mike and Eddie turned town a prestigious job.
    - 23% of their friends thought that Eddie was crazy.
    - 89% of their friends thought that Mike was crazy.
11. - Townspeople were asked about the weather.
    - 10% of people said that this winter was cruel.
    - 84% of people said that last winter was cruel.
12. - Animal control agents were surveyed on animal abuses.
    - 10% of agents said that yelling at an animal is cruel.
    - 84% of agents said that kicking an animal is cruel.
13. - Many fighting techniques can be fatal, particularly if you're on the receiving end.
    - 56% of martial artists claim that Tiger Style can be deadly to experience.
    - 78% of martial artists claim that Snake Style can be deadly to experience.

14. - Doctors were asked about animal attacks.
    - 56% of doctors felt that bee stings are deadly.
    - 78% of doctors felt that shark bites are deadly.
15. - A literature class read two different poets.
    - 57% of students in the class said that Shelley is deep.
    - 89% of students in the class said that Keats is deep.
16. - Oceanographers were consulted about the physical dimensions of different bodies of water.
    - 57% of oceanographers referred to the Caribbean as deep.
    - 89% of oceanographers referred to The Mediterranean as deep.
17. - Professors were questioned about messages in the media.
    - 25% of professors said that advertising promises are empty.
    - 60% of professors said that political promises are empty.
18. - City-dwellers were asked about urban nightlife.
    - 25% of people said that the discotheques were empty.
    - 60% of people said that the pubs were empty.
19. - Celebrities were rated on physical fitness.
    - 12% of respondents felt that Hugh Hefner is fit to run a marathon.
    - 87% of respondents felt that Lance Armstrong is fit to run a marathon.
20. - Celebrities were rated on physical fitness.
    - 12% of respondents felt that Hugh Hefner is fit.
    - 87% of respondents felt that Lance Armstrong is fit.
21. - Human rights activists were asked about political rights in America.
    - 21% claimed that Native Americans were free.
    - 91% claimed that white Americans were free.
22. - Game preserve employees were asked about two tigers.
    - 11% claim that the chained tiger is free to roam as he pleases.
    - 87% claim that the unchained tiger is free to roam as he pleases.
23. - A prisoner and a rat were stuck in a house
    - 21% of guards thought that the prisoner was free.
    - 91% of prison guards thought the rat was free.
24. - Game preserve employees were asked about two tigers.
    - 11% claimed that the chained tiger was free.
    - 87% claimed that the unchained tiger was free.
25. - Computer programmers were surveyed about user-accessibility.
    - 15% of programmers said that PCs are friendly.
    - 80% of programmers said that Macs are friendly.
26. - Dog owners were asked about unique breeds.
    - 15% of respondents said that Pit Bulls are friendly.
    - 80% of respondents said that Golden Retrievers are friendly.
27. - A group of friends was asked about two acquaintances.
    - 13% of people said that Colleen is gloomy.
    - 76% of people said that Brian is gloomy.

28. - People were asked about the weather.
    - 13% of people said that summer skies are gloomy.
    - 76% of people said that winter skies are gloomy.
29. - Marines and Army Rangers were asked about an old mission.
    - The recollections of 13% of Marines were hazy.
    - The recollections of 56% of Rangers were hazy.
30. - The Environmental Protection Agency released new findings.
    - The air in 13% of Midwestern cities is hazy.
    - The air in 56% of eastern cities is hazy.
31. - Baseball fans were asked about teams competing for the championship.
    - 48% of fans felt the Yankees are hungry.
    - 65% of fans felt the Red Sox are hungry.
32. - The man and the boy hadn't eaten in awhile.
    - 18% of nurses thought the man was hungry to eat the vegetables.
    - 58% of nurses thought the boy was hungry to eat the vegetables.
33. - Judges at an eating contest were queried.
    - 48% of judges thought George was hungry.
    - 65% of judges thought that Fred was hungry.
34. - The man and the boy hadn't eaten in awhile.
    - 18% of nurses thought that the man was hungry.
    - 58% of nurses thought that the boy was hungry.
35. - Superstitious people were surveyed.
    - 28% of respondents felt that it was lucky to find a penny.
    - 99% of respondents felt that it was lucky to find a four-leafed clover.
36. - Lottery players were asked about people who win lottery prizes.
    - 28% of players felt that scratch ticket winners are lucky.
    - 99% of players felt that Super Lotto winners are lucky.
37. - Grateful Dead fans were interviewed about drug use.
    - 12% of fans said that amphetamines can be mellow to use.
    - 89% of fans said that marijuana can be mellow to use.
38. - Grateful Dead fans were interviewed about drug use.
    - 12% of fans said that amphetamines are mellow.
    - 89% of fans said that marijuana is mellow.
39. - Military executives were asked about morale.
    - 54% of executives claimed that sailors are proud to fight for their country.
    - 78% of executives claimed that Marines are proud to fight for their country.
40. - Military executives were asked about morale.
    - 54% of executives claimed that sailors are proud.
    - 78% of executives claimed that Marines are proud.
41. - A country tried to decide whether to stick with fascism or change to democracy.
    - 11% of senators thought that creating democracy was risky.
    - 98% of senators thought that maintaining fascism was risky.
42. - Street performers were asked about their acts.
    - 34% of performers said that it is risky to juggle knives.
    - 64% of performers said that it is risky to juggle chainsaws.

43. - Parents were trying to decide which activities were more dangerous.
    - 11% of parents thought that swimming was risky.
    - 98% of parents thought that fighting was risky.
44. - Street performers were asked about their acts.
    - 34% of performers said that juggling knives is risky.
    - 64% of performers said that juggling chainsaws is risky.
45. - Sailors on a boat were asked about their officers' senses of humor.
    - 29% claimed that the skipper's jokes were a bit salty.
    - 54% claimed that the navigator's jokes were a bit salty.
46. - A family had a picnic at the beach.
    - 30% of the family thought that the food was salty to taste.
    - 85% of the family thought that the water was salty to taste.
47. - Food critics were asked to rate some dishes.
    - 29% said that the fish was salty.
    - 54% said that The veal was salty.
48. - A family had a picnic at the beach.
    - 30% of the family thought that the food was salty.
    - 85% of the family thought that the water was salty.
49. - Tourists were asked about California towns.
    - 15% of tourists said that San Diego is sleepy.
    - 80% of tourists said that Ramona is sleepy.
50. - Two boys were napping in class.
    - 15% of students said that Tom was sleepy.
    - 80% of students said that Mario was sleepy.
51. - We watched a movie and then read the book.
    - 56% of people thought that the movie was sober.
    - 92% of people thought that the book was sober.
52. - Two guys came out of a bar.
    - 56% of bar patrons thought that Dan was sober.
    - 92% of witnesses thought that Jeremy was sober.
53. - Two boys broke a window while playing stickball.
    - 12% of witnesses thought that Brian was sorry to have broken the window.
    - 67% of witnesses thought that Teddy was sorry to have broken the window.
54. - Two boys broke a window while playing stickball.
    - 12% of witnesses thought that Brian was sorry.
    - 67% of witnesses thought that Teddy was sorry.
55. - A group of friends was having a conversation.
    - 23% of the friends thought that the topic of affirmative action was thorny.
    - 78% of the friends thought that the topic of abortion was thorny.
56. - The garden was being replanted.
    - 23% of gardeners thought the tree was thorny.
    - 78% of gardeners thought that bush was thorny.
57. - Two men were fired at work.
    - 34% of coworkers said that Jenkins was unhappy to be fired.
    - 60% of coworkers said that Smith was unhappy to be fired.

58. - Two men were fired at work.
    - 34% of coworkers said that Jenkins was unhappy.
    - 60% of coworkers said that Smith was unhappy.
59. - Testers for an outdoor magazine rated ski jackets.
    - 68% of testers said that the North Face jacket was warm to wear.
    - 91% of testers said that The Patagonia jacket was warm to wear.
60. - Testers for an outdoor magazine rated ski jackets.
    - 68% of testers said that the North Face jacket was warm.
    - 91% of testers said that The Patagonia jacket was warm.

Appendix D: Experiment 5 Adjectives

Monosyllabic				Multisyllabic			
High Frequency		Low Frequency		High Frequency		Low Frequency	
Concrete	Abstract	Concrete	Abstract	Concrete	Abstract	Concrete	Abstract
bright	fast	blind	dense	beautiful	afraid	empty	distinct
dead	free	deep	neat	circular	easy	narrow	elegant
long	mean	lean	proud	level	happy	orderly	foolish
high	near	sick	weak	open	impossible	shiny	nervous
round	nice	smooth	wise	quiet	stupid	ugly	repulsive

References

Aarts, B. (2004). Modelling linguistic gradience. *Studies in Language, 28*, 1-49.

Ackerman, F., & LeSourd, P. (1997). Toward a lexical representation of phrasal predicates. In A. Alsina, J. Bresnan & P. Sells (Eds.), *Complex Predicates* (pp. 67-106). Stanford, CA: CSLI.

Ackerman, F., & Malouf, R. (2007). Parts and wholes: Patterns of relatedness in complex morphological systems and why they matter. University of California, San Diego.

Ackerman, F., & Stump, G. T. (2004). Paradigms and periphrastic expression: a study in realization-based lexicalism. In L. Sadler & A. Spencer (Eds.), *Projecting Morphology* (pp. 111-157). Stanford, CA: CSLI.

Ackerman, F., & Webelhuth, G. (1998). *A Theory of Predicates*. Stanford, CA: CSLI Publications.

Albright, A., & Hayes, B. (2002). Modeling English past tense intuitions with Minimal Generalization. In M. Maxwell (Ed.), *Proceedings of the 6th Meeting of the ACL Special Interest Group in Computational Phonology*. New Brunswick, NJ: Association for Computational Linguistics.

Albright, A., & Hayes, B. (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition, 90*, 119-161.

Anderson, N. H. (2001). *Empirical directions in design and analysis*. Mahwah, New Jersey: Lawrence Earlbaum Associates.

Andrews, A. D. (1990). Unification and morphological blocking. *Natural Language & Linguistic Theory, 8*(4), 507-557.

Aronoff, M. (1976). *Word Formation in Generative Grammar*. Cambridge, MA: MIT Press.

Baayen, R. H., Feldman, L. B., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language, 55*, 290-313.

Baayen, R. H., McQueen, J. M., Dijkstra, T., & Schreuder, R. (2003). Frequency effects in regular inflectional morphology: Revisiting Dutch plurals. In *Morphological Structure in Language Processing* (pp. 355–390). Berlin: Mouton de Gruyter.

Baayen, R. H., & Moscoso del Prado Martín, F. (2005). Semantic density and past-tense formation in three Germanic languages. *Language, 81*, 666-698.

Baayen, R. H., Piepenbrock, R., & Gulikens, L. (1995). *The CELEX lexical database (release 2), CD-ROM*. Philadelphia: Linguistic Data Consortium.

Bickel, B., Banjade, G., Gaenszle, M., Lieven, E., Paudyal, N. P., Rai, I. P., et al. (2007). Free prefix ordering in Chintang. *Language, 83*, 43-73.

Blevins, J. (2001). *Nhanda: An Aboriginal Language of Western Australia*. Honolulu: University of Hawai'i Press.

Blevins, J. P. (2005). Word-based declensions in Estonian. In G. Booij & J. Van Marle (Eds.), *Yearbook of Morphology* (pp. 1-25). Dordrecht: Springer.

Blevins, J. P. (2006). Word-based morphology. *Journal of Linguistics, 42*, 531-573.

Bloomfield, L. (1949). A set of postulates for the science of Language. *International Journal of American Linguistics, 15*, 195-202.

Bochner, H. (1993). *Simplicity in Generative Morphology*. New York: Mouton De Gruyter.

Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General, 129*, 177–192.

Boersma, P. (1997). How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam, 21*, 43-58.

Boersma, P., & Hayes, B. (2001). Empirical tests of the gradual learning algorithm. *Linguistic Inquiry, 32*, 45-86.

Boersma, P., & Levelt, C. (2000). Gradual constraint-ranking learning algorithm predicts acquisition order. In E. V. Clark (Ed.), *The Proceedings of the 30th Annual Child Language Research Forum* (pp. 229-237). Stanford, CA: CSLI.

Boyd, J. (2006). *Two stages of overgeneralization in English comparative acquisition*. Paper presented at the Boston University Conference on Language Development.

Boyd, J., & Ackerman, F. (2007). *A word-based critique of Embick (2007)*.Unpublished manuscript, University of California, San Diego.

Brassil, D. (2001). *Periphrasis, standard lexicalism and realizational lexicalism*.Unpublished manuscript, La Jolla, CA.

Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language, 34*, 383–398.

Bresnan, J. (1973). The syntax of the comparative clause construction in English. *Linguistic Inquiry, 4*, 275-343.

Bresnan, J. (2006). Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In S. Featherston & W. Sternefeld (Eds.), *Roots: Linguistics in Search of its Evidential Base*. Berlin: Mouton de Gruyter.

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation. In G. Bouma, I. Kraemer & J. Zwarts (Eds.), *Cognitive Foundations of Interpretation*. Amsterdam: Royal Netherlands Academy of Science.

British National Corpus, version 2 (BNC World). (2001). Distributed by Oxford University Computing Services on behalf of the BNC Consortium.

Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science*, *27*, 843-873.

Carstairs-McCarthy, A. (2005). Basic terminology. In P. Stekauer & R. Lieber (Eds.), *The Handbook of Word-Formation* (pp. 5-23). Dordrecht: Springer.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review, 113*, 234-272.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Clahsen, H., & Almazan, M. (1998). Syntax and morphology in Williams syndrome. *Cognition*, *68*(3), 167-198.

Clahsen, H., & Almazan, M. (2001). Compounding and inflection in language impairment: Evidence from Williams Syndrome (and SLI). *Lingua, 111*, 729-757.

Clahsen, H., & Temple, C. (2003). Words and rules in children with Williams syndrome. In Y. Levy & J. Schaeffer (Eds.), *Language competence across populations: Toward a definition of specific language impairment* (pp. 323-352). Mahwah, NJ: Lawrence Erlbaum Associates.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, *25*, 257-271.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *33*, 497-505.

Culicover, P. W., & Jackendoff, R. (2005). *Simpler Syntax*. Oxford: Oxford University Press.

Curme, G. O. (1931). *A Grammar of the English Language. Volume II: Syntax*. Essex, CT: Verbatim.

Díaz, V. G. (2006). The origin of English periphrastic comparatives. *English Studies, 87*, 707-739.

Embick, D. (2007). Blocking effects and analytic/synthetic alternations. *Natural Language and Linguistic Theory, 25*, 1-37.

Embick, D., & Noyer, R. (2007). Distributed Morphology and the syntax-morphology interface. In G. Ramchand & C. Reiss (Eds.), *The Oxford Handbook of Linguistic Interfaces* (pp. 289-324). Oxford: Oxford University Press.

Ernestus, M., & Baayen, H. (2004). Analogical effects in regular past tense production in Dutch. *Linguistics, 42*, 873-903.

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

Ferreira, V. S. (1996). Is it better to give than to donate? Syntactic flexibility in language production. *Journal of Memory and Language*, *35*, 724-755.

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, *40*, 296-340.

Ferreira, V. S., & Firato, C. E. (2002). Proactive interference effects on sentence production. *Psychonomic Bulletin & Review, 9*, 795-800.

Ferreira, V. S., Slevc, L. R., & Rogers, E. S. (2005). How do speakers avoid ambiguous linguistic expressions? *Cognition*, *96*, 263-284.

Francis, W. N., & Kucera, H. (1982). *Frequency Analysis of English Usage: Lexicon and Grammar*. Boston: Houghton Miflin.

Frank, M. (1972). *Modern English: A Practical Reference Guide*. Englewood Cliffs, NJ: Prentice-Hall.

Frege, G. (1960). Frege on definitions--I (P. T. Geach, Trans.). In P. T. Geach & M. Black (Eds.), *Translations from the Philosophical Writings of Gottlob Frege* (pp. 159-181). Oxford: Blackwell.

Gathercole, V. C. (1985). More and more and more about more. *Journal of Experimental Child Psychology, 40*(1), 73-104.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita & W. O'Neil (Eds.), *Image, Language, Brain* (pp. 95–126). Cambridge, MA: MIT Press.

Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Goldberg, A. E. (2006). *Constructions at work : the nature of generalization in language.* Oxford ; New York: Oxford University Press.

Graziano-King, J. (2003). English comparatives: a preliminary investigation. Kingsborough Community College, CUNY.

Graziano-King, J., & Cairns, H. S. (2005). Acquisition of English comparative adjectives. *Journal of Child Language, 32*(2), 345-373.

Grice, P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Gurevich, O. I. (2006). *Constructional Morphology: The Georgian Version.* University of California, Berkeley.

Hale, J. (2001). A probabilistic early parser as a psycholinguistic model. *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics,* 1-8.

Halle, M. (1962). Phonology in generative grammar. *Word, 18*, 54-72.

Halle, M., & Marantz, A. (1994). Some key features of distributed morphology. *MIT working papers in linguistics, 21*, 275-288.

Hawkins, J. A. (2000). The relative order of prepositional phrases in English: Going beyond manner-place-time. *Language Variation and Change, 11*, 231-266.

Hawkins, J. A. (2003). Efficiency and complexity in grammars: Three general principles. In J. Moore & M. Polinsky (Eds.), *The Nature of Explanation in Linguistic Theory* (pp. 121-152). Stanford, CA: CSLI Publications.

Hay, J. (2001). Lexical frequency in morphology: is everything relative? *Linguistics, 39*, 1041-1070.

Hay, J., & Baayen, R. H. (2001). Parsing and productivity. In G. E. Booij & J. van Marle (Eds.), *Yearbook of Morphology* (pp. 203-235). Dordrecht: Springer.

Hay, J., & Baayen, R. H. (2005). Shifting paradigms: gradient structure in morphology. *Trends in Cognitive Sciences, 9*(7), 342-348.

Hayes, B., & Abad, M. (1989). Reduplication and syllabification in Ilokano. *Lingua, 77*, 331-374.

Haywood, S. L., Pickering, M. J., & Branigan, H. P. (2005). Do speakers avoid ambiguities during dialogue? *Psychological Science, 16*, 362-366.

Huttenlocher, J., Vasilyeva, M., & Shimpi, P. (2004). Syntactic priming in young children. *Journal of Memory and Language, 50*, 182-195.

Jackendoff, R. (2003). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.

Jaeger, T. F. (2006). *Redundancy and Syntactic Reduction in Spontaneous Speech.* Stanford University, Stanford, CA.

Janssen, U., & Penke, M. (2002). How are inflectional affixes organized in the mental lexicon?: Evidence from the investigation of agreement errors in agrammatic aphasics. *Brain and Language, 81*, 180-191.

Jespersen, O. (1949). *A Modern English Grammar on Historical Principles, Part VII: Syntax*. Copenhagen: Ejnar Munksgaard, George Allen and Unwin.

Kager, R. (1999). *Optimality Theory*. Cambridge, MA: Blackwell.

Kay, P., & Fillmore, C. J. (1999). Grammatical constructions and linguistic generalizations: The *What's X doing Y?* construction. *Language, 75*, 1-33.

Keller, F. (2000). *Gradience in grammar: experimental and computational aspects of degrees of grammaticality.* University of Edinburgh, Edinburgh.

Keller, F., & Alexopoulou, T. (2001). Phonology competes with syntax: experimental evidence for the interaction of word order and accent placement in the realization of Information Structure. *Cognition, 79*, 301-372.

Keysar, B., Barr, D. J., & Horton, W. S. (1998). The egocentric basis of language use: Insights from a processing approach. *Current Directions in Psychological Science, 7*, 46-50.

Kiparsky, P. (2004). Blocking and periphrasis in inflectional paradigms. In G. Booij & J. Marle (Eds.), *Yearbook of Morphology* (pp. 113-135). Dordrecht: Springer.

Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research, 29*, 627-645.

Kostić, A., Marković, T., & Baucal, A. (2003). Inflectional morphology and word meaning: Orthogonal or co-implicative domains? In R. H. Baayen & R. Schreuder (Eds.), *Morphological Structure in Language Processing* (pp. 1–44). Berlin: Mouton de Gruyter.

Kraljic, T., & Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology, 50*, 194-231.

Kytö, M., & Romaine, S. (1997). Competing forms of adjective comparison in English: what could be more quicker and easier and more effective? In T. Nevalainen & L. Kahlas-Tarkka (Eds.), *To Explain the Present: Studies in the Changing English Language in Honour of Matti Rissanen* (pp. 329-352). Helsinki: Société néophilologique.

Kytö, M., & Romaine, S. (2000). Adjective comparison and standardisation processes in American and British English from 1620 to the present. In L. Wright (Ed.), *The Development of Standard English 1300-1800* (pp. 171-194). Cambridge: Cambridge University Press.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.

Lane, L. W., Groisman, M., & Ferreira, V. S. (2006). Don't talk about pink elephants! Speakers' control over leaking private information during language production. *Psychological Science, 17*, 273-277.

Langacker, R. W. (1987). *Foundations of Cognitive Grammar: Theoretical Prerequisites* (Vol. 1). Stanford, CA: Stanford University Press.

Levinson, S. C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.

Levinson, S. C. (2000). *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge, MA: MIT Press.

Levy, R. (in press). Expectation-based syntactic comprehension. *Cognition*.

Luutonen, J. (1997). *The Variation of Morpheme Order in Mari Declension*. Helsinki: Suomalais-Ugrilainen Seura.

Manova, S. (2006). *Paradigm linkage and Bulgarian verb inflection*. Paper presented at the First Conference of the Slavic Linguistics Society.

Manova, S. (2007). Personal communications, June 8-27.

Matthews, P. H. (1981). *Inflectional Morphology: A Theoretical Study Based on Aspects of Latin Verb Conjugation*. Cambridge: Cambridge University Press.

Matthews, P. H. (1991). *Morphology* (2nd ed.). Cambridge, MA: Cambridge University Press.

McCarthy, J., & Prince, A. (1994). The emergence of the unmarked: Optimality in prosodic morphology. *Proceedings of the North East Linguistic Society, 24*(33-379).

McCawley, J. D. (1988). *The Syntactic Phenomena of English* (Vol. 2). Chicago: University of Chicago Press.

McFarland, T. (2007). *Free affix order in Filomeno Mata Totonac*. Paper presented at the Annual Meeting of the Linguistic Society of America.

Mehler, J., Sebastian, N., Altmann, G., Dupoux, E., Christophe, A., & Pallier, C. (1993). Understanding compressed sentences: The role of rhythm and meaning. *Annals of the New York Academy of Sciences, 682*, 272-282.

Milin, P., Đurđević, D. F., & Moscoso del Prado Martín, F. (2007). The psychological reality of inflectional paradigms. University of Novi Sad.

Mondorf, B. (2002). The effect of prepositional complements on the choice of synthetic or analytic comparatives. In H. Cuyckens & G. Radden (Eds.), *Perspectives on Prepositions* (pp. 65-78). Tubingen: Niemeyer.

Mondorf, B. (2003). Support for more-support. In G. Rohdenburg & B. Mondorf (Eds.), *Determinants of Grammatical Variation in English* (pp. 251-304). Berlin: Mouton de Gruyter.

Moscoso del Prado Martín, F., Baayen, R. H., & Kostić, A. (2004). Putting the bits together: an information theoretical perspective on morphological processing. *Cognition, 94*, 1-18.

Pinker, S. (1999). *Words and Rules: The Ingredients of Language*. New York: Perennial.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 28,* 73-193.

Pinker, S., & Prince, A. (1994). Regular and irregular morphology and the psychological status of rules in grammar. In S. D. Lima, R. L. Corrigan & G. K. Iverson (Eds.), *The Reality of Linguistic Rules* (pp. 321-351). Philadelphia: Johns Benjamins.

Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition. *Cognition, 38,* 43-102.

Plunkett, K., & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition, 48*, 21-69.

Poser, W. J. (1992). Blocking of phrasal constructions by lexical items. In I. Sag & A. Szabolsci (Eds.), *Lexical Matters* (pp. 111-130). Stanford, CA: CSLI.

Prince, A., & Smolensky, P. (1993). Optimality Theory: Constraint interaction in generative grammar. Ms., Rutgers University and the University of Colorado.

Ramscar, M. (2002). The role of meaning in inflection: Why the past tense does not require a rule. *Cognitive Psychology, 45*(1), 45-94.

Rohdenburg, G. (1996). Cognitive complexity and increased grammatical explicitness in English. *Cognitive Linguistics, 7*, 149–182.

Roland, D., Elman, J. L., & Ferreira, V. S. (2006). Why is *that*? Structural prediction and ambiguity resolution in a very large corpus of English sentences. *Cognition, 98*, 245-272.

Rosch, E. (1973). Natural categories. *Cognitive Psychology, 4*, 328-350.

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Vol. 2, pp. 216-271). Cambridge, MA: MIT Press.

Sadler, L., & Spencer, A. (2000). Syntax as an exponent of morphological features. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology* (pp. 71-96). Dordrecht: Kluwer.

Salminen, T. (1997). *Tundra Nenets inflection*. Helsinki: Suomalais-Ugrilainen Seura.

Sapir, E. (1921). *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace.

Savage, C., Lieven, E., Theakston, A., & Tomasello, M. (2003). Testing the abstractness of children's linguistic representations: lexical and structural priming of syntactic constructions in young children. *Developmental Science, 6*, 557-567.

Savage, C., Lieven, E., Theakston, A., & Tomasello, M. (2006). Structural priming as implicit learning in language acquisition. *Language Learning and Development*, 27-49.

Sims, A. (2005). Avoidance strategies and periphrasis: evidence against blocking analyses. Ohio State University.

Slevc, L. R. (2007). *Saying what's on your mind: Working memory effects on syntactic production.* University of California, San Diego, La Jolla.

Sorace, A., & Keller, F. (2005). Gradience in linguistic data. *Lingua, 115*, 1497-1524.

Spencer, A. (2004). Morphology - an overview of central concepts. In L. Sadler & A. Spencer (Eds.), *Projecting Morphology* (pp. 67-109). Stanford, CA: CSLI.

Stevens, A. M. (1971). Fixed morpheme order. *Linguistic Inquiry, 2,* 420-421.

Stewart, T., & Stump, G. (2007). Paradigm Function Morphology and the morphology-syntax interface. In G. Ramchand & C. Reiss (Eds.), *The Oxford Handbook of Linguistic Interfaces* (pp. 383-421). Oxford: Oxford University Press.

Stump, G. T. (2001). *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.

Stump, G. T. (2002). *Morphological blocking and Panini's principle*. Paper presented at the UCSD Periphrasis and Paradigms Workshop.

Temperley, D. (2003). Ambiguity avoidance in English relative clauses. *Language, 79*, 464-484.

Tomasello, M. (2000). The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences, 4,* 156-163.

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.

Wasow, T. (1997). End-weight from the speaker's perspective. *Journal of Psycholinguistic Research, 26*, 347-361.

Watters, D. E. (2006). Notes on Kusunda Grammar: A Language Isolate of Nepal. *Himalayan Linguistics Archive, 3,* 1-182.

Wheeldon, L. R., & Levelt, W. J. M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language, 34,* 311-334.

Wittgenstein, L. (1968). *Philosophical Investigations* (G. E. M. Anscombe, Trans. third ed.). Oxford: Blackwell.