

©Copyright 2014

Thomas Trimble



# Adjectives in the LinGO Grammar Matrix

Thomas Trimble

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

Master of Science

University of Washington

2014

Reading Committee:

Emily M. Bender, Chair

Barbara Citko

Program Authorized to Offer Degree:  
Linguistics



University of Washington

**Abstract**

Adjectives in the LinGO Grammar Matrix

Thomas Trimble

Chair of the Supervisory Committee:  
Associate Professor Emily M. Bender  
Department of Linguistics

The LinGO Grammar Matrix (Bender et al. 2002, 2010) provides a system for user-linguists to jump start the creation of starter Head-driven Phrase Structure Grammar precision grammars (Pollard and Sag 1994), with semantic representations in Minimal Recursion Semantics (Copestake et al. 2005). The Grammar Matrix provides an online questionnaire for users to describe their target language in a user-friendly and typologically motivated fashion. This description is utilized to produce customized, language-specific rule definitions extending a core, near universal set of types available to any grammar.

I propose and implement a new library for intersective adjectives cross-linguistically, considering both attributive and predicative constructions, editing and extending the core grammar while adding additional capabilities to the online customization system to analyze adjectives in target languages and generate language-specific customized grammars with analyses of adjectives. I present a broad typological review the behavior of adjectives, including the morphology and syntax of adjectives, along with an overview of the literature on the semantics of adjectives. I also present a review of the adjectives in several large implemented deep linguistic HPSG grammars in the DELPH-IN formalism.

I develop a cross-linguistic analysis of adjectives, adapting previous DELPH-IN analyses to cover significant amounts of new data. The analysis relies not only on definitions in the lexicon, but also on defining the syntactic behavior of adjectives in the morphology. I present a computational implementation of this analysis as an extension to the Grammar Matrix. Finally, I present an



evaluation of this extension, showing that the extension achieves 100% coverage of development language test suites and 100% coverage of held out test language test suites, with minimal spurious ambiguity.





## TABLE OF CONTENTS

|                                                                             | Page |
|-----------------------------------------------------------------------------|------|
| Glossary . . . . .                                                          | iv   |
| Chapter 1: Introduction . . . . .                                           | 1    |
| Chapter 2: Adjectives in the World's Languages . . . . .                    | 3    |
| 2.1 Some Terminology . . . . .                                              | 5    |
| 2.2 What is an adjective? . . . . .                                         | 6    |
| 2.2.1 Word Order, Semantic Features, Agreement, and Incorporation . . . . . | 7    |
| 2.2.2 Syntactic Constructions of Adjectives . . . . .                       | 8    |
| 2.2.3 Semantics of Adjectives . . . . .                                     | 9    |
| 2.2.4 Argument Structure of Adjectives . . . . .                            | 10   |
| 2.3 Adjectives in Attributive Constructions . . . . .                       | 11   |
| 2.3.1 Word Order . . . . .                                                  | 12   |
| 2.3.2 Agreement . . . . .                                                   | 14   |
| 2.3.3 Incorporated Adjectives . . . . .                                     | 15   |
| 2.3.4 Multiple Adjectives . . . . .                                         | 16   |
| 2.4 Adjectives in Predicative Constructions . . . . .                       | 18   |
| 2.4.1 Copula Complement vs. Stative Predicative . . . . .                   | 19   |
| 2.4.2 Mixed Cases . . . . .                                                 | 22   |
| 2.4.3 Morphology . . . . .                                                  | 25   |
| 2.4.4 Copulas . . . . .                                                     | 27   |
| 2.5 Summary of Phenomena . . . . .                                          | 28   |
| Chapter 3: Theoretical Background . . . . .                                 | 30   |
| 3.1 The LinGO Grammar Matrix . . . . .                                      | 30   |
| 3.2 HPSG and MRS . . . . .                                                  | 31   |
| 3.2.1 HPSG . . . . .                                                        | 32   |
| 3.2.2 MRS . . . . .                                                         | 41   |
| 3.2.3 Summary . . . . .                                                     | 43   |

|            |                                                                                |     |
|------------|--------------------------------------------------------------------------------|-----|
| 3.3        | Adjectives in HPSG Grammars . . . . .                                          | 44  |
| 3.3.1      | ERG . . . . .                                                                  | 44  |
| 3.3.2      | Jacy . . . . .                                                                 | 49  |
| 3.3.3      | gCLIMB . . . . .                                                               | 51  |
| 3.3.4      | The Grammar Matrix & Language CoLLAGE . . . . .                                | 54  |
| 3.4        | Summary . . . . .                                                              | 56  |
| Chapter 4: | Analysis . . . . .                                                             | 58  |
| 4.1        | Basic Adjectives in the Grammar Matrix . . . . .                               | 59  |
| 4.1.1      | A Cross-linguistic Analysis of Adjectives . . . . .                            | 60  |
| 4.1.2      | Language-Specific Constraints . . . . .                                        | 67  |
| 4.1.3      | Summary of Adjective Analysis . . . . .                                        | 70  |
| 4.2        | Copulas in the Grammar Matrix . . . . .                                        | 71  |
| 4.2.1      | Basic Copula Definition . . . . .                                              | 71  |
| 4.2.2      | Copula Inflection and Agreement . . . . .                                      | 72  |
| 4.2.3      | Copula Complement Selection . . . . .                                          | 73  |
| 4.2.4      | Summary . . . . .                                                              | 74  |
| 4.3        | Mixed Type Languages . . . . .                                                 | 74  |
| 4.3.1      | Optional Inflection . . . . .                                                  | 74  |
| 4.3.2      | Optionally Copula Complement . . . . .                                         | 75  |
| 4.4        | Constrained Argument Agreement . . . . .                                       | 76  |
| 4.4.1      | Lexical or Morphological Process? . . . . .                                    | 77  |
| 4.4.2      | Morphological Analysis of Constrained Argument Agreement . . . . .             | 77  |
| 4.5        | Morphologically Incorporated Adjectives . . . . .                              | 79  |
| 4.6        | Summary . . . . .                                                              | 82  |
| Chapter 5: | Implementation . . . . .                                                       | 84  |
| 5.1        | Extending the Customization System User Interface . . . . .                    | 85  |
| 5.1.1      | Lexicon . . . . .                                                              | 86  |
| 5.1.2      | Morphology . . . . .                                                           | 91  |
| 5.1.3      | Switching . . . . .                                                            | 92  |
| 5.1.4      | Summary . . . . .                                                              | 98  |
| 5.2        | The Lexical and Morphological Components of the Customization System . . . . . | 98  |
| 5.2.1      | Lexicon . . . . .                                                              | 100 |
| 5.2.2      | Morphology . . . . .                                                           | 103 |
| 5.2.3      | Summary . . . . .                                                              | 106 |

|            |                                                 |     |
|------------|-------------------------------------------------|-----|
| 5.3        | Constraining the Customization System . . . . . | 107 |
| 5.3.1      | Encoding Linguistic Generalizations . . . . .   | 107 |
| 5.3.2      | Reducing User Errors . . . . .                  | 107 |
| 5.4        | Summary . . . . .                               | 108 |
| Chapter 6: | Evaluation . . . . .                            | 109 |
| 6.1        | Pseudo-Languages & Regression Tests . . . . .   | 109 |
| 6.1.1      | Pseudo-Languages . . . . .                      | 109 |
| 6.1.2      | Summary . . . . .                               | 117 |
| 6.2        | Natural Languages . . . . .                     | 118 |
| 6.2.1      | Illustrative Languages . . . . .                | 119 |
| 6.2.2      | Held Out Languages . . . . .                    | 125 |
| 6.3        | Summary . . . . .                               | 133 |
| Chapter 7: | Conclusion & Future Work . . . . .              | 134 |
| 7.1        | Summary . . . . .                               | 134 |
| 7.2        | Future Work . . . . .                           | 135 |
| 7.3        | Closing . . . . .                               | 136 |

## GLOSSARY

1/2/3: first/second/third

SG/PL: singular/plural

MASC/FEM: masculine/feminine

NA/NI: animate/inanimate

NOM/ACC/DAT: nominative, accusative, dative case

AGT/PAT: agentative, patientative case

DEF/INDEF: definite/indefinite

PRES/PAST/FUT: present/past/future tense

SUBJ/OBJ/COMP: subject/object/complement

AUX: auxiliary

COP: copula

DET: determiner

PRON: pronoun

REL: (semantic) relation *or* relative clause

POL: polite

LIT.: literally

## ACKNOWLEDGMENTS

First, I owe all thanks to Mom, from whom I learned too many things to count and not enough. I miss you every day. I owe so much to my wife Cailin, for being my best friend and for supporting me every day, emotionally, spiritually, and in every way.

To Emily Bender, for introducing me to implemented linguistics, helping me refine how to be a good scientist, and reminding me to look at the data! To Barbara Citko, for introducing me to linguistics, her great contributions to this work, and always being willing to help out and chat linguistics.

To all my many linguistics professors for their inspiration and guidance: to Edith Aldridge for inspiration in syntax, guiding me towards implemented linguistics, and mentoring me so excellently; to Toshi Ogihara for helping me to expand my linguistic horizons and reminding me to laugh; to Fei Xia for grounding my academic work in reality; to Gina Levow for helping me to dream big; and to Lee Osterhout for giving me the opportunity to explore.

To all my great peers and fellow DELPH-IN students: to Sanghoun Song for being a great friend and mentor; to Michael Goodman for helping me along and always smiling; to Joshua Crowgey for getting me started and starting with me; and to Woodley Packard for laughing about words and giving me too much of his time.

To my many great mentors: to Shane Gordon for inspiring me to lead and biting off as much as I could; to Stephen Bonfoey for helping me to learn about myself and others; to Cassandra Bethard for mentoring me to grow and think strategically; to Andrew Kersten for leading me through hard times.

To my many great friends: to Ross Henderson for growing in academia together and always willing to chat about language though we always disagree; to Ben Burkhalter for being fantastic and a huge support; to Jared Kramer for struggling through together and succeeding together.

Finally to my family: to Debra for always being there; to Craig for always helping out; to David for the fun times and laughter; and to my nieces, aunts, uncles, and many cousins because family.

Lastly, thank you to Dad and Mom for which I owe everything.

## **DEDICATION**

In memory of my mother, Marta Lynn McAlexander Trimble, for infinite courage and inspiration.





## Chapter 1

### INTRODUCTION

The LinGO Grammar Matrix (Bender et al. 2002, 2010) provides a system for user-linguists to jump start the creation of starter HPSG precision grammars (Head-driven Phrase Structure Grammar, Pollard and Sag 1994). Using the Grammar Matrix customization system (Drellishak 2009; Bender et al. 2010), user-linguists choose options fitting their target language from a typologically informed questionnaire. The customization system automatically generates a customized machine-readable HPSG starter grammar from an array of stored analyses to cover various phenomena in the target language described by the user-linguist. Grammar definition files are encoded in the Type Description Language, or TDL, which closely mirrors HPSG type definitions (Copestake 2002). This grammar is downloadable and functions with an array of DELPH-IN processing tools.<sup>1</sup> The output grammar includes core definitions of near-universal phenomena (such as head-adjunct syntactic rules and morphologically affixing lexical rules) as well as language-specific phenomena encoded across several files, including those dedicated to syntax, morphology, and the lexicon. The definitions to elicit and generate the language-specific constraints are stored in server-side libraries, which, when combined with the user-linguist's answers to the questionnaire, generate the language-specific HPSG definitions.

Grammar Matrix grammars can be used in conjunction with existing tools to parse input strings into syntactic and semantic analyses, as well as generate natural language strings from input semantic representations. This way, grammatical hypotheses can be codified, implemented, and tested. Tools for test suite management, treebanking, and other applications have been implemented within the DELPH-IN code-base.<sup>2</sup>

I focus on developing a novel library for defining adjectives in the Grammar Matrix customization system. My goals for this project are the following:

---

<sup>1</sup><http://www.delph-in.net>

<sup>2</sup>More applications at: <http://moin.delph-in.net/DelphinApplications>

- Describe the range of variation of adjectives across the world's languages to define the scope of phenomena to be implemented.
- Develop HPSG analyses of the syntax and semantics of these phenomena.
- Encode the analyzed phenomena into set of user interface options for user-linguists to select in the Grammar Matrix customization system.
- Develop a set of tests to evaluate the analyses and prevent regressions in later system development.

To do this, I survey the relevant typological literature to define the scope of my work within the range of variation in the world's languages. In order to capture the relevant range of the morphological, syntactic, and semantic behavior of intersective adjectives, I survey adjectives in attributive and predicative constructions as well as copulas supporting predicative adjective constructions (such as the English verb *to be* in *the dog is big*). On the basis of this survey, I develop HPSG analyses to include in the Grammar Matrix. Finally, grammars and test suites for various natural languages as well as abstract pseudo-languages designed to represent specific phenomena are developed and implemented to test the analyses.

Chapter 2 presents my survey of the typological variation of adjectives, including a cross-linguistic discussion of word categories, adjective characteristics, and syntactic and semantic behaviors of attributive and predicative adjectives. This discussion provides proper foundation for the analysis of adjectives. Chapter 3 provides an overview of the theoretical background: HPSG and MRS, and discusses the HPSG analysis of adjectives and copulas in other DELPH-IN grammars, including the ERG, Jacy, and gCLIMB grammars. Section 3.3.4 provides an overview of the LinGO Grammar Matrix customization system. Chapter 4 presents my analysis of the phenomena implemented in the Grammar Matrix. Chapter 5 details my implementation of adjectives in the Grammar Matrix customization system and user interaction. Chapter 6 presents an evaluation of the system, including coverage over illustrative languages analyzed during development and coverage over held-out languages tested after implementation was complete. Chapter 7 provides conclusions, reflection, and directions for future work.

## Chapter 2

### ADJECTIVES IN THE WORLD'S LANGUAGES

The Grammar Matrix is a tool to aide grammar engineering. Goals of grammar engineering include developing analyses of phenomena in languages to model morphological patterns, syntactic structure, compositional semantics, and grammaticality. To provide a foundation to an analysis of these phenomena as they relate to adjectives, this chapter provides a broad overview of linguistic variation of adjectives in the world's languages with a focus for developing a foundation for precise grammatical description across as many languages as possible. To do this, I review typological literature regarding adjectives.

Typologists generally agree that all languages have at least two open and universal syntactic and semantic word classes: nouns and verbs (Dixon 2004). The status of further word classes is not solidified (see Dixon 1982, Bhat 1994, Wetzer 1992, Wetzer 1996, and Baker 2003). This leads typologists generally to define adjectives by a semantic notion, though there is sometimes disagreement (for further discussion, see Dryer 2007b; Dixon 2004; Schachter and Shopen 2007). Abstracted slightly from Schachter and Shopen 2007, I define adjectives as syntactically and morphologically distinct words that specify a property of an entity.

Adjectives have several distinct semantic functions, which are sometimes analyzed as set relations between the adjective and the specified entity (Partee 1995; Flickinger et al. 2003).

In DELPH-IN grammars, there is a distinction between scopal and non-scopal adjectives, where scopal adjectives include *tough* in the English *it's a tough problem to solve* (Flickinger 2000; Siegel and Bender 2002). Scopal adjectives do not directly specify a property of a noun, but rather a scopal adjective's external semantic argument is the scopal argument of its complement so that the adjective can also interact with scopal ambiguities with respect to quantifiers in the MRS semantic representation. The scopal argument of the complement is typically an infinitival verb in English, such as the above with *tough* or *easy* in *Kim is easy to please* (Flickinger 2000; Flickinger et al. 2003).

Non-scopal adjectives include those that Partee 1995 calls intersective, subsective, and non-intersective. These include adjectives like *big* in the English *the black dog barked*, which combine with heads resulting in a semantic notion whose interpretation can be modeled as referencing the intersection of the sets denoted by the noun and the adjective. For instance, many adjectives in English are intersective adjectives, where the resulting phrase refers to an intersection between the set of entities that are *big* and the set of entities that are *dogs*. While both scopal and non-scopal adjectives have been analyzed in HPSG and implemented in DELPH-IN grammars, my focus is on non-scopal adjectives.<sup>1</sup>

Adjectives generally come in two syntactic constructions, attributive and predicative. Attributive constructions are those in which the adjective is a modifier of the noun, either as an adjunct or an incorporated affix. Predicative constructions are those in which the adjective is associated directly with the subject of the clause, broadly distinguished into copula complement adjectives, such as those in English, and intransitive stative predicates, such as those in Mupun (Frajzyngier 1993, from Dryer 2007a).

- (1) a. He found the **red** ball (Attributive)
- b. The ball is **red** (Predicative)

The methodology of this chapter differs slightly from other typological surveys of adjectives and copulas. Stassen 2003, 2013 focus on comparing adjectives to nouns and verbs, searching for similarities, especially in languages where adjectives pattern very similarly to nouns or verbs. Alternatively, my focus is in building a set of distinguishing factors that can be encoded into user interface options in the customization system. Consequently, I focus on what distinguishes adjectives from other word classes, as opposed to how they might be similar.

The goal of this chapter is to provide a broad review the syntax, morphology, and semantics of intersective adjectives cross-linguistically, with a focus on developing a set of choices to elicit input from the user-linguist of the Grammar Matrix customization system in producing language-specific analyses. Section 2.1 provides a brief overview of the terminology I use throughout the entirety

---

<sup>1</sup>Following the ERG (Flickinger 2000), I give all non-scopal adjectives the same MRS representation in my analysis.

of this document. Section 2.2 further discusses the attributes of adjectives and distinguishes them from other word categories. Section 2.3 discusses adjectives in attributive constructions, focusing on variations in word order, agreement, and briefly discussing adjective incorporation. Section 2.4 discusses stative predicate constructions, copula complement constructions, mixed cases (where one language exhibits both of the previous items), the morphology of predicative constructions, and a brief overview of copulas. Section 2.5 provides an overview of the phenomena and conclusions.

## **2.1 Some Terminology**

Adjectives are distinguished from other lexical categories by how they behave. In the HPSG literature, and more specifically in the DELPH-IN style grammars, this terminology is encoded into features of analyses (Pollard and Sag 1994; Flickinger 2000; Bender et al. 2002). The specifics of this will be discussed thoroughly in Chapters 3 and 4. This section briefly establishes some terminology used throughout this chapter and the rest of the document.

First is **lexical category**. Lexical categories are a much studied topic in linguistic literature (Dixon 1982, Bhat 1994, Wetzer 1992, and Wetzer 1996). I define lexical categories as morphologically, syntactically, and semantically distinct classes of elements of the lexicon.

Second is **semantic features**. These come in two pertinent varieties: features of events and features of entities. Semantic features of events include tense, aspect, and mood, and are usually associated with verb-like lexical categories. Semantic features of entities include person, number, and gender, and are usually associated with noun-like lexical categories.

Third is **inflection**. Inflection is the morphological manifestation of semantic features of events and entities. This is in contrast to the fourth term, **agreement**, which is the morphological manifestation of features of heads on which dependents rely. Typically, agreement only occurs in semantic features of entities, and semantic features of entities often only appear morphologically as agreement morphology (such as definiteness in many Indo-European languages).

Fifth is **particle**. For the purposes of this document, I define particles as typically semantically empty lexemes that depend on their head whose primary function is to manifest a feature of their head, similar to inflectional and agreement morphology.

Last is **pivots**. I define pivots here informally as a term to refer to representative features or

distinctions to be used in categorizing and differentiating linguistic behaviors. These pivots are used when defining the user interface of the questionnaire.

## 2.2 What is an adjective?

Adjectives specify a property of an a noun. While most languages are analyzed with a lexical category of adjective, some languages are less appropriately analyzed this way, including those relying on noun-noun compounds or other techniques (Schachter and Shopen 2007). Some languages, such as English, have large, morphologically, syntactically, and semantically distinct classes of adjectives, while others, such as Objibwa, have adjectives which are less morphologically or syntactically distinct from nouns or verbs (Stassen 2013). Other languages have small, closed-class sets of adjectives, such as Igbo (Emenanjo 1978). Some languages have adjectives that can be said to be modifying other adjectives, such as in English: *the spicy hot chili* or *the dark blue light*, where the first adjective could be said to be modifying the property of the second, as opposed to the noun. These examples are considered outside my scope.

Adjectives in languages tend towards a prototypical set of relations: DIMENSION, COLOR, AGE, and VALUE, such as the antonymic pairs in English *long/short*, *black/white*, *new/old*, and *good/bad* (Dixon 1977, 2004). Some languages, such as Igbo, do not have adjectives representing relations beyond these whatsoever (Emenanjo 1978). The attested extent of Igbo adjectives are shown in (2). Other semantic concepts expressed as adjectives in many Indo-European languages are expressed either through analytic or synthetic combinations of existing adjectives, or through other means.

(2) All attested adjectives in Igbo (Emenanjo 1978)

| Dimension    | Color        | Age         | Value      |
|--------------|--------------|-------------|------------|
| ukwu 'large' | ojii 'black' | ọhụrụ 'new' | ọma 'good' |
| nta 'small'  | ọca 'white'  | ocye 'old'  | ọjọọ 'bad' |

From the perspective of extending the Grammar Matrix customization system, I leave the lexical status of a particular lexeme as a verb, noun, or adjective as undetermined. Instead, I assume that the type “adjective” is a universally relevant type to the work of a grammar engineer, regardless of the target language. While adjectives are incorporated into the “language independent” core Grammar

Matrix file (and have been since near its inception, Bender et al. 2002), grammarians are not required to extend these types or implement instances of them in their grammar.

### 2.2.1 *Word Order, Semantic Features, Agreement, and Incorporation*

Adjectives are distinguished from other lexical categories in four primary ways. First is word order, where the adjective may appear as an adjunct to the noun, complement of a copula, or in a different configuration than other lexical categories. For instance, in French, adjectives can appear as a copula complement (like nouns) or as a noun modifier (such as nouns in a noun-noun compound):

#### (3) French

- a. je    veux       des               choux-fleurs  
      1.SG want.1.SG DET.PL.INDEF cabbage-flowers-PL  
      ‘I want some cauliflower’, lit. ‘I want some cabbage flowers’ [fra]
- b. je    veux       des               fleurs       rouges  
      1.SG want.1.SG DET.PL.INDEF flower-PL red.MASC-PL  
      ‘I want some red flowers’ [fra]
- c. ils   sont       des               fleurs  
      3.PL COP.3.PL DET.PL.INDEF flower-PL  
      ‘they are flowers’ [fra]
- d. ils   sont       rouges  
      3.PL COP.3.PL red.MASC-PL  
      ‘they are red’ [fra]

However, adjectives in French are distinguished from nouns in several ways. The second distinguishing factor is agreement: first, adjectives agree with the noun they modify or take as their subject in number and gender. Second, most adjectives in French appear post head, while noun-noun compounds are headed similarly to English (Dixon 2004). Third, nouns in French require determiners, while adjectives forbid them. Consequently, it seems apt to analyze languages such as French to have a distinguished word class of adjectives from nouns that require different treatments of word order, agreement, and other features.

Adjectives in languages such as Ojibwa, where adjectives pattern in a more verb-like way, inflect for semantic features, like tense/aspect/mood, and have similar negation strategies and complementation patterns. Stassen 2013 details how in various languages, these factors differ between verbs

and adjectives. For instance, within a given language, the negation strategy for verbs might differ from the negation strategy for adjectives (such as in Gumbaynggir (Eades 1979, pp. 332)).

Another key source of variation cross-linguistically is whether the adjective is a separate word or an affix to the modified noun. The latter appears in polysynthetic languages (and possibly others), such as the Algonquian languages Abenaki and Penobscot. Note that in (4), the Penobscot adjectives with translate to *bad* and *red* appear as affixes to the noun they modify.

(4) Penobscot

- a. wənɪhləkən          mətahsəm  
     wə-nəh-l-ək<sup>w</sup>-əne mat-ahsəm  
     3-kill-RP-INV-N bad-dog  
     ‘the bad dog kills him’ [aaq-pen] (Quinn 2006)
- b. mk<sup>w</sup>əhpske  
     mhk<sup>w</sup>-hpsk-e-w  
     red-rock-LV<sup>NI</sup>-w  
     ‘it is a red rock’ [aaq-pen] (Quinn 2006)

Understanding these differences and the behavior of adjectives cross-linguistically is key to extending the language independent core grammar of the Grammar Matrix to include these sorts of adjectival constructions for use by user-linguists. The following sections will further discuss these differences in an attempt to capture generalizations of how languages differ, and what sort of pivots might be developed to capture these variations.

While adjectives are primarily distinguished by word order, agreement, and inflection, other features can also distinguish adjectives from other word categories. Specifically, adjectives also vary from verbs or nouns in their word order in non-matrix clauses, negation constructions, and surely more. This is somewhat problematic in that several of these other phenomena have not yet been analyzed for inclusion in the Grammar Matrix. For the purposes of the Grammar Matrix, these sorts of variations are left for future work on the word order, constituent negation, and other libraries.

### 2.2.2 Syntactic Constructions of Adjectives

While many adjectives in many of the world’s languages can appear in either attributive or predicative constructions, there are adjectives which can only appear in one or the other. Consider these



examples from English:

(5) English Attributive *vs.* Predicative (Flickinger 2000)

- a. the **big** dog barks
- b. the dog is **big**
- c. the **mere** thought startled me
- d. \* the thought is **mere**
- e. \* the **awake** dog barks
- f. the dog is **awake**

While most adjectives in English (and many other languages) appear in both syntactic constructions, there are many adjectives that do not. There are three possibilities: attributive-only, predicative-only, and those that appear in both, which I will refer to as **regular**.

Examples of attributive-only adjectives in English (from the ERG (Flickinger 2000)) include *folk*, *mere*, *mock*, *mid*, *other*, *overall*, *pseudo*, and *utter*. Examples of predicative-only adjectives in English include many of the *a-* affixed words, such as *abed*, *ablaze*, *awake*, etc., as well as others such as *gone* and *lengthwise*. Words in other languages have been analyzed as attributive or predicative only, such as Japanese's *tannaru* ('mere') being analyzed as an attributive only adjective (Siegel and Bender 2002).

### 2.2.3 Semantics of Adjectives

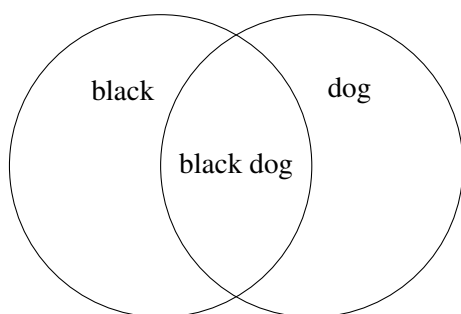
This section is intended to provide a broad and shallow summary of the semantics of adjectives as described in the semantics literature and existing DELPH-IN grammars. The primary purpose is to specify and constrain the target of investigation. I make no attempt at any novel understanding of the semantics of adjectives.

I see two broad categories of adjectives. First is termed scopal adjectives, where the semantic head of the phrase is the adjective's complement, such as *tough* in the phrase *it is a tough problem to solve*. The adjective is the semantic head of the phrase, allowing for quantifiers to scope between the adjective and the noun. I leave scopal adjectives to future work.

Second is termed here as non-scopal adjectives, adjectives which Partee 1995 describes as words which add greater specificity to the meaning of nouns they predicate, further distinguishing be-

tween **intersective**, **non-intersective**, and **subsective** adjectives (see also Vendler 1968; Larson 1995; Cinque 2010).

First is the intersective adjective, where the resulting noun phrase refers to the intersection of the set of things described by the noun and the set of things described by the adjective. Therefore, the *black dogs* in *black dogs bark* refers to intersection of the set of *black* things and the set of *dogs*.



The latter two semantic types of adjectives include adjectives such as *skillful* and *fake*, these being subsective and non-intersective, respectively. This is due to constructions such as:

Subsective and non-intersective adjectives (adapted from Kennedy 2012)

(6) the skillful violinist plays well

(7) the fake gun broke

In (6), it is not the case that the violinist is skillful at everything, such as being a surgeon, but rather at being a violinist. In the case of (7), a fake gun is not a gun at all, the referent is not an intersection of *fake* and *gun* entities, nor is it a subset of *gun* entities that are *fake*. I set these questions of subsective and non-intersective adjectives aside and constrain my focus to intersective adjectives.

#### 2.2.4 Argument Structure of Adjectives

While adjectives are generally intransitive in that they combine strictly with one argument, either subject or modificand, there are many examples of transitive adjectives. Transitive adjectives include those like *worth* in English (Maling 1983), as in the sentence *it is worth my time*. Some scopal adjectives are also transitive, such as *tough* in *it's a tough problem to solve*. These questions of transitivity in adjectives are considered beyond my scope and left to future work.

### 2.3 Adjectives in Attributive Constructions

Adjectives appear as syntactic modifiers of nouns consistently across many languages. In other languages, similar semantics are expressed through adjectives incorporated into the noun. These attributive adjectives vary in word order and morphology. The following is a set of examples of adjectives in attributive constructions:

(8) Adjectives in attributive constructions (adjectives in **bold**)

a. English

The **big** dog barks [eng]

b. Apatani

aki **kinyo** píne

dog small bark.PST

‘the small dog barks’ [apt] (adapted from Abraham 1985)

c. French

la **petite** femme rit

DET.FEM.DEF small-FEM woman laugh.3.SG.PRES

‘the small woman laughs’ [fra]

d. Penobscot

wəníhləkon mətahsəm

wə-nəh-l-ək<sup>w</sup>-əne **mat**-ahsəm

3-kill-RP-INV-N bad-dog

‘the bad dog kills him’ [aaq-pen] (Quinn 2006)

While the vast majority languages have attributive adjectives, Dryer 2013 survey of 1366 languages found that 0.3% of the languages in the sample do not seem to exhibit attributive adjectives. Instead, adjectives appear as predicatives in relative clauses:

(9) Diegueño (Langdon 1970)

a. iikwich=ve=ch aq ku-nemshap=vu aakwal

man=DEF=SUBJ bone REL.SUBJ-white=DEF lick

‘The man licked the white bone’ [dih]

b. kwenychekwii=ve=ch hekwany ku-mii=vu selyewelyuu

old.woman=DEF=SUBJ baby REL.SUBJ-cry=DEF tickle

‘The old woman tickled the baby that cried’ [dih]

Key cross-linguistic variations of the syntax of attributive adjectives include variations in word order, agreement, and morphological status. These are discussed in more detail in the follow sections.

### 2.3.1 *Word Order*

Languages vary in the order of noun and adjective. Dryer 2013 makes four primary distinctions, which follow the logical possibilities:

1. adjective-noun
2. noun-adjective
3. neither dominant
4. no attributive adjectives

Dryer 2013's survey of word order in 1366 languages found that 64% of languages are noun-adjective, 27% of languages are adjective-noun, 8% are either position, and only 0.3% of languages were found to not have attributive adjectives.

However, these classifications were made language-wide. In terms of grammar engineering, it is less interesting to make a broad statement about a language. Precision judgments of given utterances are more interesting. Subsequently, it can be of more interest to make specific analyses for different adjectives or different forms of a given adjective. For instance, French can be analyzed to have three classes of adjectives: most are post-head, some are pre-head, and some are either (Dixon 2004):

(10) French (adapted from Dixon 2004; Waugh 1977)

- a. le                    chien rouge      aboie  
    DET.MASC.DEF dog   red.MASC bark-3.SG  
    'the red dog barks' [fra]
- b. \* le                    rouge      chien aboie  
    DET.MASC.DEF red.MASC dog   bark-3.SG  
    'the red dog barks' [fra]
- c. le                    petit            chien aboie  
    DET.MASC.DEF small.MASC dog   bark-3.SG  
    'the small dog barks' [fra]

- d. \* le chien petit aboie  
 DET.MASC.DEF dog small.MASC bark-3.SG  
 ‘the small dog barks’ [fra]
- e. le chien grand aboie  
 DET.MASC.DEF dog big.MASC bark-3.SG  
 ‘the large dog barks’ [fra]
- f. le grand chien aboie  
 DET.MASC.DEF big.MASC dog bark-3.SG  
 ‘the great dog barks’ [fra]

The distinction between pre-head *grand* and post-head *grand* is that the post-head version has a meaning similar to “large” or “tall” while the pre-head version is ambiguous between this same meaning as the post-head variety and a meaning similar to “great” or “strong” in English (Vaughan 1977). This sort of semantic change is termed by Stassen 2013 as stage versus individual predicates (Carlson 1977a,b), where the stage level predicate is interpreted to be a temporal or temporary attribute, such as greatness after a success, whereas individual level predicates are interpreted to be inherent properties, such as physical size.

Because adjectives differ within languages, it is interesting to look at behaviors of specific adjectives. To abstract from Dryer 2013’s work, possible positions of a given class of adjectives within a language are:

1. adjective-noun
2. noun-adjective
3. either position
4. some other position

Therefore, these distinctions can properly capture languages with varying strategies. The *some other position* option is provided in the hypothetical case that adjectives appear in a position other than the other three. The purpose of this option is exclusively to inform users targeting languages with this behavior that it is not supported (see §5.1.1 for further details).

### 2.3.2 Agreement

In many languages, attributive adjectives agree with several grammatical features of the nouns they modify, including person, number, gender, case, and more. Adjectives in French agree in gender and number, while adjectives in German agree in case, gender, definiteness, and what is commonly referred to as “strength” (Rankin and Wells 2010; Fokkens 2011):

(11) a. German

Boris hat den wissenschaftlern einen  
 Boris has DET.PL.ACC.DEF scientist.MASC.DAT DET.SG.MASC.ACC.INDEF  
 großen Auftrag gegeben  
 big-MASC.ACC assignment.MASC.ACC give.PST

‘Boris gave the scientists a big assignment’ [deu] (Adapted from McFadden 2006)

b. den wissenschaftlern ist ein  
 DET.PL.ACC.DEF scientist.PL.DAT COP.3.SG.PRES DET.SG.MASC.NOM  
 großer auftrag gegeben worden  
 big-MASC.NOM assignment.NOM give.PST become  
 ‘the scientists were given a big assignment’ [deu] (Adapted from McFadden 2006)

c. eine starke frau geht  
 DET.FEM.NOM.INDEF strong-FEM.NOM woman.FEM goes  
 ‘a strong woman walks’ [deu] (Zeller 2003)

d. ein starker mann geht  
 DET.MASC.NOM.INDEF strong-MASC.NOM man.MASC goes  
 ‘a strong man walks’ [deu] (Zeller 2003)

e. ein rotes buch fiel  
 DET.MASC.NOM.INDEF red-MASC.INDEF book.NOM fall.PST  
 ‘a red book fell’ [deu] (adapted from above and Lee-Schoenfeld 2007<sup>2</sup>)

f. das rote buch fiel  
 DET.MASC.NOM.DEF red-MASC.DEF book.NOM fall.PST  
 ‘the red book fell’ [deu] (adapted from above and Lee-Schoenfeld 2007)

While adjectives in German agree with the noun they modify, they do not agree with their subjects in predicative constructions.

---

<sup>2</sup>These examples were also confirmed by proficient German speaker Marni Wiebe.

## (12) German

- a. Hans glaubt daß er krank ist  
 Hans think-PRES that 3.SG ill COP.3.SG.PRES  
 ‘Hans thinks that he is ill’ [deu] (Adapted from Jary 2004; Richter 2006)
- b. die musik ist laut  
 DET.FEM.NOM music.FEM COP.3.SG.PRES loud  
 ‘the music is loud’ [deu] (Adapted from Jary 2004; Richter 2006)
- c. ihr ist kalt  
 3.SG.FEM COP.3.SG.PRES cold  
 ‘she is cold’ [deu] (Zeller 2003)
- d. ich bin groß  
 1.SG COP.1.SG.PRES big  
 ‘I am tall’ [deu] (Landman and Morzycki 2002)

Lastly, some languages’ adjectives also have a particle that agrees in definiteness. For instance, Greek:

## (13) Greek Adjectival Definiteness Markers

- a. to vivlio to kokkino to megalo ...  
 DEF book DEF red DEF big ...  
 ‘the big red book ...’ [ell] (Alexiadou and Wilder 1998)
- b. to megalo to vivlio to kokkino ...  
 DEF red DEF big DEF book ...  
 ‘the big red book ...’ [ell] (Alexiadou and Wilder 1998)

2.3.3 *Incorporated Adjectives*

Some languages, including polysynthetic languages such as the Algonquian languages of northeastern North America, have noun affixes which contribute a prototypically adjectival semantic relation to the sentence.

## (14) Penobscot

- a. wəníhləkon mətahsəm  
 wə-nəh-l-ək<sup>w</sup>-əne mat-ahsəm  
 3-kill-RP-INV-N bad-dog  
 ‘the bad dog kills him’ [aaq-pen] (Quinn 2006)

These affixes are termed to be **incorporated stems**, following Baker 1985.

### 2.3.4 Multiple Adjectives

Languages such as English and other Indo-European languages tend to allow the use of multiple adjectives to modify a single modificand:

(15) Multiple adjectives

a. English

the big black dog barks [eng]

b. German

die                großen                schwarzen                hund bellt  
DET.FEM.NOM big-NOM.MASC black.NOM.MASC dog bark-PST

‘the big black dog barked’ [deu] (adapted from Hankamer and Lee-Schoenfeld 2005)

However, there are exceptions and limitations to this. Maori forbids multiple adjective constructions, using conjoined noun phrases in its place:

(16) Maori attributive adjectives

a. ka    taka    ōna    whatu    ki    runga    i    tētahi                tangata    tino    nui    tangata    tino  
TAM fall   his   eye    to top    at DET.INDEF man    very big man    very  
mōmona  
fat

‘his eyes lit on a very big very fat man’ [mri] (adapted from Bauer et al. 1997)

b. \* ka    taka    ōna    whatu    ki    runga    i    tētahi                tangata    tino    nui    tangata  
TAM fall   his   eye    to top    at DET.INDEF man    very big very  
mōmona  
fat

‘his eyes lit on a very big very fat man’ [mri] (adapted from Bauer et al. 1997)

It has also been noted that while incorporated adjectives, such as those in the Algic language family, can be added to a single stem, there is at least a discourse limit of two to three, at which point additional adjectives appear as relative clauses (Quinn, p.c.).<sup>3</sup>

---

<sup>3</sup>This is similar to a hypothetical discourse limit in English, where strings like *the big, loud, red, long, hairy, round, strong dog barks* are marked, while strings like *the big, loud, red dog who was also long, (and) hairy, (and) round, and strong barks*, are much less so, especially in speech.



It is also of interest to note languages such as French, where different adjective classes have different directions of modification and can still apply together:

(17) French multiple adjectives

- a. le                      petit                      chien   noir                      aboie  
 DET.DEF.MASC.SG small.MASC.SG dog.SG black.MASC.SG bark-3.SG.PRES

‘the small black dog barks’ [fra]

One last area of brief discussion is on the topic of ordering restrictions between multiple adjectives (see Sproat and Shih 1991; Cinque 2010; Laenzlinger 2005). This phenomena is exemplified in data such as the following:

(18) Multiple adjective ordering restrictions

- a. the small green Chinese vase broke [eng] (Sproat and Shih 1991, pp. 565)
- b. ?the green Chinese small vase broke [eng] (Sproat and Shih 1991, pp. 565)
- c. lǜ    xiǎo   huāpíng  
 small green vase  
 ‘small green vase’ [cmn] (Sproat and Shih 1991, pp. 566)
- d. \* xiǎo   lǜ    huāpíng  
 green small vase  
 intended: ‘green small vase’ [cmn] (Sproat and Shih 1991, pp. 566)
- e. la   sola   possibile   invasione   romana   della   Tracia  
 the only possible   invasion   Roman of-the Thrace  
 ‘the only possible Roman invasion of Thrace’ [ita] (Cinque 2010, pp. 1)
- f. la   sola   invasione   possibile   della   Tracia  
 the only invasion   possible   of-the Thrace  
 ‘the only possible invasion of Thrace’ [ita] (Cinque 2010, pp. 1)
- g. \* la   sola   invasione   possibile   romana   della   Tracia  
 the only invasion   possible   Roman of-the Thrace  
 intended: ‘the only possible Roman invasion of Thrace’ [ita] (Cinque 2010, pp. 1)

While these data show interesting ordering constraints, I am not convinced of their syntactic nature. For instance, given the proper discourse scenarios, *the green Chinese small vase broke* is a viable English sentence. While a further study of these phenomena is warranted, I leave these to future work.

## 2.4 *Adjectives in Predicative Constructions*

Predicative constructions, those in which the adjectives serve as the semantic (and often times syntactic) head of the clause, generally fall into two syntactic patterns: adjective as copula complement and adjective as intransitive predicate. For example:

### Copula complement

- (19) a. English

John is tall [eng]

- b. Irish

is            breoite é  
COP.PRES ill     he

‘He is ill’ [gle] (Greene 1966)

### Intransitive predicate

- (20) a. Tagalog

mapanganib ang lumapit sa ahas  
dangerous NOM ACT.approach DAT snake

‘To go near a snake is dangerous’ [tgl] (Schachter 1987, from Falk 2006, pp.

57)

- b. Lakota

ma-si’ca  
1.SG.PAT-bad

‘I am bad’ [lkt] (Mallinson and Blake 1981, from Falk 2006, pp. 11)

- c. Bororo

i-kure-re  
1.SG-tall-NEUTRAL

‘I am tall’ [bor] (Crowell 1979)

Note that while the literature refers to the direct combinations as intransitive predicates, I will refer to these as stative predicates.

However, some languages only have adjectives in attributive constructions, such as Hua, where

the adjective must be in an attributive construction, and what is expressed in other languages is expressed as an attributive form on a predicative noun:

(21) Hua

- a. Bura fu nupa fu baie  
that pig black pig COP  
'That pig is a black pig' [ygr] (Haiman 1978)
- b. \*Bura fu nupa baie  
that pig black COP  
'That pig is a black' [ygr] (Haiman 1978)

In the copula complement constructions, a semantically empty copula is used to connect the adjective to its subject (Flickinger 2000; Pustet 2003). Much of the work in the typological literature (e.g. Schachter and Shopen 2007; Dryer 2007b,a; Stassen 2003, 2013) focuses on how predicative adjectives differ from verbs or predicative noun constructions. Stassen 2013 identifies three primary distinguishing factors between what is termed *verbal* and *nonverbal* adjectives:

1. Agreement
2. Copula
3. Negation

While agreement and copulas are core to adjectival constructions in languages that use them (i.e. it is necessary to analyze copulas in order to analyze adjectives in these languages), negation is tangential in that not all adjectival constructions are negated. Subsequently, I focus on agreement and copula constructions.

This section is organized into four subsections: first is discussion of copula complement and stative predicate adjectives, second is further discussion of mixed cases, third is discussion of inflection and agreement on stative predicates, and last is discussion of copulas.

#### 2.4.1 Copula Complement vs. Stative Predicative

The primary differences between predicative adjectives across the world's languages are whether the adjective may appear as a copula complement or not, and what sort of inflection it has. Stassen 2013 terms these 'copula complement' and 'intransitive predicate' adjectives. I will refer to Stassen

2013's 'intransitive predicates' as **stative predicates** because not all predicative adjectives are intransitive. This section discusses each of these in turn.

### *Copula Complement*

Many languages, including many Indo-European languages, utilize a semantically empty copula to connect an adjective to its subject:

(22) Irish

- a. is                breoite é  
COP.PRES ill        3.SG  
'He is ill' [gle] (Greene 1966)
- b. Tá                sé        mór  
COP.PRES 3.SG big  
'He is big' [gle] (Carnie 1995)
- c. Bhí                sé        cliste  
COP.PST 3.SG clever  
'he was clever' [gle] (Carnie 1995)

(23) Spanish

- a. la                idea es                buena  
DET.DEF.FEM idea COP.3.SG.PRES good-FEM  
'The idea is good' [spa] (Gallego 2006)
- b. ella                es                alta  
3.SG.FEM COP.3.SG.PRES tall-FEM  
'She is tall' [spa] (Bleam 2000)

(24) Arabic<sup>4</sup>

- a. kan        l-kalb        kabir  
COP.PST DEF-dog big  
'the dog was big' [arb] (Attia et al. 2008)
- b. al-kalb        kabir  
DEF-dog big  
'the dog is big' [arb] (Attia et al. 2008)

These sorts of constructions have three important features. The first important feature is the agreement and inflection patterns. In languages with copulas, the tense, aspect, and mood inflection tends to be on the copula, as opposed to the adjective. However, agreement can still occur between

---

<sup>4</sup>Thanks for Clara Gordon for help with these examples

the adjective and the subject. For instance, in Spanish, the copula agrees with the subject in person and number while the adjective agrees with the subject in gender; the copula also inflects for tense.

The second important feature is whether the copula that combines with the adjective is also used for other non-verbal predicates. Some languages have multiple copulas that take different sorts of complements, where one will take an adjective and another a noun, for instance. Pustet 2003 identifies several languages with copulas which select for a particular complement type, such as Bambara [bam], Epena Pedee [sja], Shilluk [shk], Chalcatongo Mixtec [mig], and Thai [tha]. For instance, Epena Pedee:

(25) Epena Pedee

- a. úsa p<sup>h</sup>aimáa pa-hí  
dog black COP-PST  
'it was a black dog' [sja] (Harms 1994, from Pustet 2003)
- b. pía bi  
good COP.PRES  
'that is good' [sja] (Harms 1994, from Pustet 2003)

The third important feature is the order of copula and complement. In some languages, the copula complement comes after the copula, such as English or French. In other languages, such as Miyako [mvi] of Japan, the adjective comes before the copula.

*Stative Predicate*

Adjectives as stative predicates, those which combine with their subject directly and inflect for tense, aspect, mood, agreement, or some collection of these, are common outside of the Indo-European language family (Stassen 2013). Some examples follow:

(26) a. Tiwi

tungk Waltiringa pumpuka  
stringy.bark good  
'the stringy bark is good' [tiw] (Osborne 1974)

b. Bororo

i-kure-re  
1.SG-tall-NEUTRAL  
'I am tall' [bor] (Crowell 1979)

In some languages, the adjective inflects for tense, aspect, or mood (such as Bororo), while in other languages, the construction has an underspecified or implicitly specified tense (such as Tiwi). For instance, in Russian, in the present tense, predicative adjectives combine directly with the subject (Dryer 2007a):

(27) Russian

- a. on                      molod-oj  
     3.SG.MASC.NOM young-MASC.SG.NOM  
     ‘he is young’ [rus] (Engelberg 2005)

This construction is sometimes analyzed as a zero copula. As the specifics of the zero copula hypothesis would require a much more thorough research of copulas in the world’s languages, I do not posit a zero category, along the lines of the surface oriented HPSG literature (Pollard and Sag 1994; Copestake 2002). Subsequently, I analyze this sort of construction as a stative predicate.

#### 2.4.2 *Mixed Cases*

Stassen 2013 defines *verbal* and *non-verbal* encoding of adjectives, using aforementioned criteria (agreement, copulas, and negation) to distinguish them. From these, Stassen 2013 describes cases where one language has both *verbal* and *non-verbal* adjectives. In the case of the Grammar Matrix, this sort of organization of the lexicon is left to the user-linguist. However, in order to enable these mixed behavior phenomena, I survey the data presented in Stassen 2013 with a goal to enable the development of analyses of these adjectival patterns. Stassen 2013 terms languages which exhibit adjectives with both *verbal* and *nonverbal* as *mixed case* languages. Stassen 2013 identifies two varieties:

1. Split: there are several classes of adjectives which behave differently, taking different inflectional patterns, appearing as a copula complement or not, etc.
2. Switching: individual adjectives inflect differently or optionally appear as a copula complement, typically with a semantic difference, commonly a stage vs. individual predicate distinction.

This phenomenon can be broken down into three intersecting phenomena.

1. Single/Multiple adjective classes

2. Optional or construction-constrained agreement
3. Adjective optionally a copula complement

To summarize Stassen 2013's account in these terms, 'split' languages have multiple adjective classes, usually with consistent feature marking and complementation patterns. 'Switching' languages instead have one adjective class with some combination of optional feature marking and adjectives optionally appearing as copula complement. While it seems possible for a language to exhibit both of these behaviors, my survey of the typological literature and language data did not uncover any such language.

These phenomena will be discussed below.

### *Optional Inflection or Particles*

Rama is a case of a split mixed case language, where a class of adjectives, including one meaning *hungry*, inflects similarly to verbs, while another class of adjectives, including one meaning *good*, takes no inflection:

- (28) Rama
- a. Verb  
m-upluui-i  
2-dream-PRES  
'You are dreaming' [rma] (Colette Grinevald, from Stassen 2013)
  - b. Stative predicate inflecting like verbs  
nsut tiiskibadut s-angaling-i  
1.PL children 1.PL-hungry-PRES  
'We children are hungry' [rma] (Colette Grinevald, from Stassen 2013)
  - c. Stative predicate without inflection  
ning suurak mliima  
this pineapple good  
'This pineapple is good' [rma] (Colette Grinevald, from Stassen 2013)

Maori is a case of a switching mixed case language, where adjectives are marked with either verbal features or nominal features, INCEP and INDEF respectively below, through their particles.

## (29) Maori

## a. Verbal predicate

ka oma te kootiro  
 INCEP run DET.DEF girl  
 ‘The girl runs’ [mri] (Biggs 1969, from Stassen 2013)

## b. Verb-like adjectival predicate

ka pai te whare nei  
 INCEP good DET.DEF house this  
 ‘This house is good’ [mri] (Biggs 1969, from Stassen 2013)

## c. Noun predicate

he kiwi teera manu  
 DET.INDEF kiwi DET.DEM bird  
 ‘This bird is a kiwi’ [mri] (Biggs 1969, from Stassen 2013)

## d. Noun-like adjectival predicate

he pai te koorero  
 INDEF good DET.DEF talk  
 ‘This talk is good’ [mri] (Biggs 1969, from Stassen 2013)

In Maori, the distinction is a semantic distinction, where the verb-like inflection indicates a stage level predicate and the noun-like inflection indicates an individual level predicate.

*Optionally Copula Complement*

The construction where an adjective optionally combines with a copula or a subject is termed here optionally copula complement. Many Indo-European languages have similar constructions, including Russian, Hebrew, Turkish, and Arabic.

## (30) a. Russian

vy sčastlivy-j  
 2.NOM.POL happy-NOM.MASC.SG  
 ‘you are happy’ [rus] (Hahm et al. 2007)

b. vy byli sčastlivy-j  
 2.NOM.POL COP.PST.PL happy-NOM.MASC.SG  
 ‘you were happy’ [rus] (Hahm et al. 2007)

These constructions often differ in semantic features. For instance, in Russian, the stative predicate construction is present tense, while the adjective is licensed as a copula complement in other



tenses.

### *Constrained Argument Agreement*

Another similar behavior to optional inflection appears in languages like German, where inflection only occurs in the attributive construction and is forbidden in the predicative construction.

(31) German

- a. der                      große                      hund bellt  
 DET.MASC.NOM.SG big-MASC.NOM dog bark-PST  
 ‘the black dog barked’ [deu] (adapted from Hankamer and Lee-Schoenfeld 2005)
- b. ich    bin                      groß  
 1.SG COP.1.SG.PRES big  
 ‘I am tall’ [deu] (adapted from Landman and Morzycki 2002)

### 2.4.3 *Morphology*

Predicative adjectives vary in their morphology across languages. The morphology of the world’s adjectives generally fall into two categories:

1. Tense, aspect, mood, and other verbal feature inflection
2. Person, number, gender, and other nominal feature agreement

These are detailed in the following subsections.

### *Inflection*

Rama is an example of a language where (some) predicative adjectives inflect for tense:

(32) Rama

- a. Tense inflection on verbs  
 m-upluui-i  
 2-dream-PRES  
 ‘You are dreaming’ [rma] (Colette Grinevald, from Stassen 2013)
- b. Tense inflection on adjectives  
 nsut tiiskibadut s-angaling-i  
 1.PL children 1.PL-hungry-PRES  
 ‘We children are hungry’ [rma] (Colette Grinevald, from Stassen 2013)

Luo is an example of a language where (some) predicative adjectives inflect for aspect (Luo may be analyzed as a switching language):

(33) Luo

- a. án      má-be'r  
 1.SG PERF-good  
 'I am good' [luo] (Tucker and Bryan 1966)

There were not any examples found in the literature of adjectives in copula complement constructions inflecting for semantic features. See §2.4.4 for more details on copula agreement.

*Agreement*

French (and other Romance languages) is an example of predicative adjectives agreeing in gender with their subject:

(34) French

- a. le                      chien est              grand  
 DET.DEF.MASC dog    COP.3.SG red.MASC  
 'the dog is red' [fra]
- b. la                      chaise est              grande  
 DET.DEF.MASC chair    COP.3.SG big-FEM  
 'the chair is big' [fra]

Russian adjectives agree in gender, number, and case with their subject (or modificand):

(35) Russian

- a. on                      molod-oj  
 3.SG.MASC.NOM young-SG.MASC.NOM  
 'he is young' [rus] (Engelberg 2005)
- b. vy                      sčastlivy-e  
 2.NOM.PL happy-NOM.PL  
 'y'all are happy' [rus] (Hahm et al. 2007)

Consequently, it appears that the features predicative adjectives might inflect for include tense, aspect, mood, or other semantic features of events, and agree with their subject in features such as person, number, gender, case, or other semantic features of entities.

#### 2.4.4 Copulas

The term copula is used to describe a diverse set of lexical categories, as discussed in Pustet 2003, including those which link subjects and predicates, identify their subject and complement, or host semantic features. Some examples of copulas taking adjectival complements include:

- (36) a. English  
the dog **is** big
- b. German (Landman and Morzycki 2002)  
ich **bin** groß  
1.SG COP.PRES.1.SG big  
'I am tall' [deu]
- c. Russian (Hahm et al. 2007)  
vy **byli** sčastlivy-j  
2.NOM.POL COP.PST.PL happy-NOM.MASC.SG  
'you were happy' [rus]

I assume the following definition of adjective-complement adjectives (from Pustet 2003):

Copula: a semantically empty 'linker' between subject and predicate.

In the tradition of HPSG, this definition is codified as joining the complement and subject via two key constraints: first, identifying the semantic index of the subject with the external argument of the complement; second, the copula does not contribute an elementary predication (see §3.2.1). While copulas also have many other interesting syntactic and semantic properties, I focus on the above definition of copulas as it applies to adjectives. Moreover, as one of my goals is to develop copulas congruent with adjectival phenomena and analyses, I primarily focus on adjectival complement copulas, as opposed to those that take nominal, verbal, or sentential complements.

Along the lines of Flickinger 2000, I analyze the copula in predicative adjective constructions as being semantically empty, in that it does not contribute a semantic predicate to the phrase (see also Pustet 2003). Note that Flickinger 2000 does distinguish between several copulas, including an identity copula in NP predicate constructions such as *Fido is a dog*, but the copula taking adjective complements is semantically empty. This analysis for copulas in predicative adjective constructions is congruous with the many languages that utilize stative predicates to express similar semantic notions without a copula, for instance with predicative adjectives or nouns. Adjective complement

copulas are raising verbs, identifying the external argument of its complement (the adjective) with the index of its subject. See § 3.2.1.

As discussed previously, copulas often inflect for tense, aspect, or mood, especially in languages where the adjective is excluded from such inflection. Indeed, Pustet 2003 describes such a function of copulas as being thought of as the core or even exclusive function of copulas. In some languages, copulas also agree with their subjects in person, number, gender, or some other lexical feature.

No discussion was discovered in the typological literature or language data of copulas agreeing with their complement.

## **2.5 Summary of Phenomena**

This survey has only scratched the surface of adjectival constructions, however, for the purpose of developing a library for adjectives in the Grammar Matrix customization system, it is necessary to focus on the primary distinctions between and within languages. This section summarizes these distinctions from the typological literature.

The first distinction is a question of whether a given adjective in a given language is attributive-only, predicative-only, or both. The languages surveyed here and in the literature seem to exhibit adjectives with both behaviors most often, however there are examples of different constraints throughout a language and within a language. English has classes of adjectives that are attributive only, predicative only, or both. Hua seems to only have attributive adjectives (Haiman 1978). Diegueño (Langdon 1970) appears to only have predicative adjectives, a relative clause being used in place of most attributive adjectives.

The second distinction is a matter of a word order and supporting words. There are separate considerations for the attributive construction as well as the predicative construction. In the attributive construction, does the adjective come before or after the modificand, or are both allowed, with or without a semantic difference? In the predicative construction, does the adjective appear as a copula complement, or not? In the case that it does, does the adjective (or more broadly, the copula complement) come before or after the copula? Is this congruous or not with auxiliary-main verb combinations?

The third distinction is morphology, which includes the expression of semantic features (such as

tense, aspect, and mood) and nominal features (such as case, person, number, and gender). Adjectives that appear as stative predicates tend to inflect for some semantic feature, but this is not always the case. In languages with copulas, the copula tends to inflect for semantic features appropriate for events. Nominal feature agreement is common in both attributive and predicative constructions, while there are sometimes differences in agreement patterns between these constructions. Lastly, in polysynthetic languages, there are adjectives that attach as affixes to noun stems.

Chapter 3 details the theoretical background and previously implemented analyses of adjectives. Chapter 4 presents an HPSG analysis of these various constructions as implemented in the Grammar Matrix.

## Chapter 3

### THEORETICAL BACKGROUND

My work is rooted in a theoretical background of HPSG, MRS, and the DELPH-IN framework. Head Driven Phrase Structure Grammar (HPSG, Pollard and Sag 1994) is a strongly lexicalist constraint based typed feature structure grammar theory. Minimal Recursion Semantics (MRS, Copestake et al. 2005) is a flat, graph-like semantic representation of elementary predications (EPs), their syntactic features, and their semantic argument structure. The DELPH-IN framework is a formalism and collection of tools within the HPSG and MRS framework, including tools for parsing, generation, treebanking, translation, and test suite management. This section discusses HPSG and MRS, as well as provides an overview of the Grammar Matrix customization system.

#### **3.1 *The LinGO Grammar Matrix***

The Grammar Matrix is a tool designed to help jump start the creation of DELPH-IN style HPSG grammars (Bender et al. 2002, 2010). The system is built around a ‘language independent’ core of near universal analyses. These analyses are extended on a language by language basis through the Grammar Matrix customization system, where user-linguists fill out a typologically grounded questionnaire about their target language. The answers to these questions are then used to produce language-specific syntax, lexicon, and morphology files, which combine with the core grammar to produce a language-specific customized grammar.

As diagrammed in Figure 3.1, the user interacts with the customization system through a web-based questionnaire. Choices are stored in a list of choice IDs (such as the name of a specific question, lexical item, etc.) and the provided answer. While the user interacts with the system, a validation step ensures that choices made are congruent with each other. For instance, if a user indicates that their language has determiners, they must define at least one determiner type and lexical item. Any errors are presented to the user as hover text.

Once the user is ready to produce a grammar from their choices, the system produces a grammar

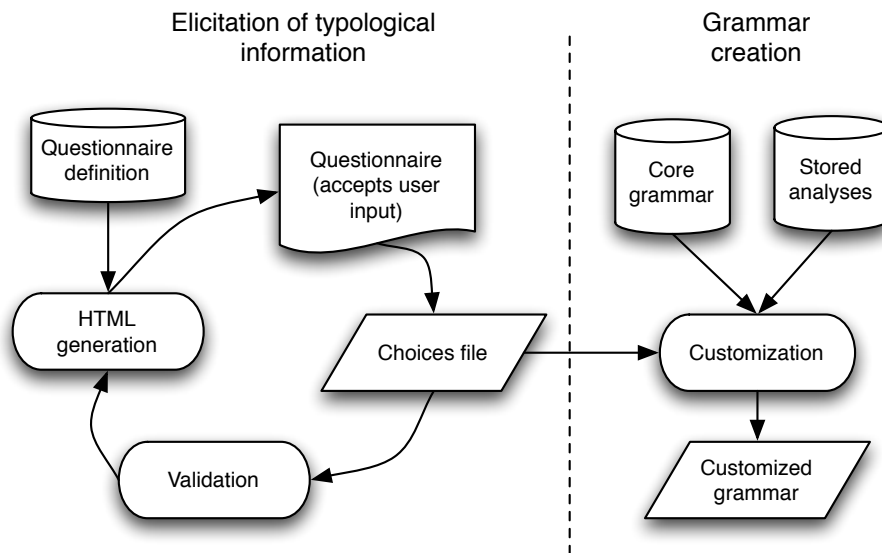


Figure 3.1: Schematic of the Grammar Matrix customization system (Bender et al. 2010)

by combining the core grammar with types and instances defined on the basis of stored analyses and user input. The result is a customized HPSG grammar which produces semantic representations in the MRS formalism.

The Grammar Matrix customization system is extensible through new libraries. Existing libraries include syntactic phenomena: word order (Drellishak 2009), sentential negation (Crowgey 2012), and argument optionality (Saleem 2010); grammatical features: person, number, gender, and case (Drellishak 2009); the lexicon: nouns, verbs, determiners, and auxiliaries (Bender et al. 2010); morphology and morphotactics (O’Hara 2008; Goodman 2013); and information structure (Song 2013). My goal is to extend the lexicon and morphology components to enable users to build analyses of adjectives for their language. The specifics of this extension is discussed in Chapter 4.

### 3.2 HPSG and MRS

The Grammar Matrix’s lexical and syntactic constraints and values are defined in HPSG, with semantic representations in MRS. §3.2.1 discusses HPSG, specifically the DELPH-IN Joint Reference Formalism implementation of HPSG, and aspects of it important to my analysis. §3.2.2 discusses MRS, the semantic representation used in DELPH-IN grammars.

### 3.2.1 HPSG

HPSG (Pollard and Sag 1994) is a unification based typed feature structure theory of grammar. An HPSG grammar is a hierarchy of types, where each type is defined as a structured set of features stored in an attribute-value matrix (AVM). The type hierarchy relies heavily on multiple inheritance, which allows a type to be defined as an extension of one or more less specific types, inheriting the constraints and definitions of each parent type. This way, the root of the hierarchy is the least specific type, a type from which all types inherit, and the leaves are the most specific. The type hierarchy defines instances of phrase structure rules, lexical rules, and lexical items, using multiple inheritance to capture generalizations.

Within the general framework of HPSG, there is variation in the specific formalisms and analyses adopted. My work is situated within the DELPH-IN Joint Reference Formalism (Copestake 2002), a specific implementation of HPSG. Additionally, my work builds on top of the Grammar Matrix core grammar, which is a specific set of types and constructs built within the Joint Reference Formalism (Bender et al. 2002). My description in this chapter is focused on the DELPH-IN Joint Reference Formalism and its implementation in the Grammar Matrix.

The key operation within HPSG is unification. There are two sorts of definitions: values and indices. Two types A & B are said to unify if:

- One or both of A and B is unspecified
- Both A and B are specified and A equals B
- Both A and B are specified and A is a subtype of B or vice versa
- Both A and B are specified and they share a common subtype
- Both A and B are specified and A and B are indices, in this case A and B are identified

Type definitions within an AVM are stored in a set of graph-like structures, including syntax and semantics (or SYNSEM), phonology, information structure, argument structure, inflection flags, and orthography. The following is an example of a truncated type definition of the adjective *big* from the English Resource Grammar (ERG, Flickinger 2000):



(37) Lexical type of *big* from the ERG

$$\left[ \begin{array}{l} \text{ORTH} \quad \langle \text{“big”} \rangle \\ \text{SYNSEM} \quad \left[ \begin{array}{l} \text{LKEYS.KEYREL.PRED “_big\_a\_1\_rel”} \\ \text{PHON.ONSET } con \end{array} \right] \end{array} \right]$$

This lexical item inherits from *aj\_-i-er\_le*, which inherits from several other types unifying into the following constraints (and more; additional constraints excluded for brevity):

(38) Type definition of adjectives in the ERG (truncated & simplified)

$$\left[ \begin{array}{l} \text{SYNSEM} \quad \left[ \begin{array}{l} \text{LOCAL} \quad \left[ \begin{array}{l} \text{CAT} \quad \left[ \begin{array}{l} \text{HEAD.MOD.LOCAL.CONT.HOOK.INDEX } \boxed{1} \\ \text{VAL.SPR.FIRST.LOCAL.CONT.HOOK.XARG } \boxed{1} \end{array} \right] \\ \text{CONT.HOOK} \quad \left[ \begin{array}{l} \text{INDEX } \boxed{1} \\ \text{XARG } \boxed{2} \end{array} \right] \end{array} \right] \\ \text{LKEYS} \quad \left[ \begin{array}{l} \text{KEYREL} \quad \left[ \begin{array}{l} \text{ARG0 } \boxed{1} \\ \text{ARG1 } \boxed{2} \end{array} \right] \end{array} \right] \end{array} \right] \end{array} \right]$$

Matching indices inside of squares are said to be **identified**, which unifies their values. This has the effect of constraining possible values and “assigning” values in a non-technical sense. Here, the function of the constraints in (38) is to constrain the type’s modificand and subject of the adjective to be identified with the type’s first semantic argument (ARG1). This results in the type’s first semantic argument being properly identified in both attributive and predicative constructions.

### *Typed feature structures and rules*

Everything in a DELPH-IN grammar is a typed feature structure, including lexemes, morphological rules, phrase structure rules, and syntactic features. New grammatical description in the form of new lexical items, lexical types, phrase structure rules, or other analysis is conducted through the formulation of new types or instances which inherit from one or more existing types. Sentences are considered well-formed when the types used to describe a string do not conflict with one another. More specifically, a string is well-formed when the types in a given utterance unify with the initial

symbol, usually defined as a sentence or fragment. Because the initial symbol is defined as a bundle of features, as opposed to an atomic Context-Free Grammar-style symbol, this is also known as the root condition.

Lexical items are stored as leaves of type hierarchies which define specific values and constraints. These include syntactic features, semantic predicates, and morphological constraints. Morphological constraints come in the form of *luk*<sup>1</sup> features which constrain whether a particular lexical item is available to be a daughter of particular syntactic rule.

Syntactic phrase structure rules take one to three daughters, which may be lexical items, the output of lexical rules, or the mother of other phrase structure rules, and return a new feature structure with the constraints of the rule and certain constraints of the lexical items. Most phrases are “headed” in that rules match a HEAD value of a phrase, the part of speech and related properties, with some other unit. For instance, the subject-head (or subj-head) phrase combines a “nouny” subject with a “verby” head, typically resulting in a type that satisfies the root condition.

### *Multiple Inheritance*

Types are defined in the DELPH-IN Joint Reference Formalism in a type hierarchy. Everything in the DELPH-IN style type hierarchy must unify with *\*top\**, a special symbol at the top of a type hierarchy. Types inherit from their supertypes and add additional constraints such that the leaves are the most specific constraints and the *\*top\** is not constrained. Types can also inherit from multiple supertypes, allowing for constraints to be collapsed into one type. These hierarchies of constraints can be visualized in figures such as Figure 3.2:

The hierarchy is designed such that the leaf values unify in place of their supertypes, so that values can be underspecified by using a less specific supertype. For instance, specifying a value as *bool* allows it to unify with either *—* or *+*, because both of these inherit from *bool*.

Similarly, types can inherit from multiple supertypes and subsequently inherit all the features of all of their supertypes. This allows for feature hierarchies to remain compact and avoid restating similar constraints.

---

<sup>1</sup>*luk*-valued features are *+*, *—*, or not applicable, along with *na-or-+* and *na-or-—*. These provide a minimal extension over boolean values.

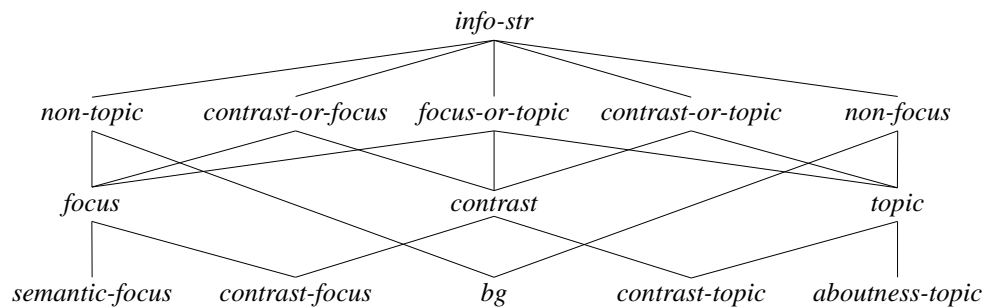
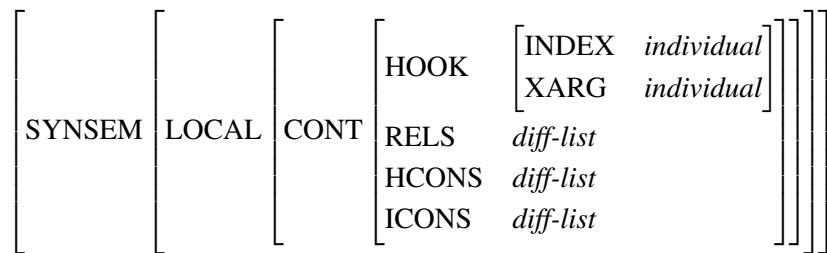


Figure 3.2: Information Structure Hierarchy exemplifying Multiple Inheritance (from Song 2013)

### Structure of a lexeme

The Grammar Matrix builds upon the English Resource Grammar’s type definitions for lexemes and many features. Semantic content is defined for a given lexeme in a feature structure *cont*, with externally visible content stored in the feature *HOOK* and a list of local EPs stored as a list of relations in the feature *RELS*:

(39) Definition of *CONT*



The type *HOOK* contains the *index* and *xarg*, representing the type’s semantic index and external argument, respectively. These are crucial to the semantics of adjectives, as discussed in §3.2.2.

Relevant syntactic constraints are defined in a feature structure *cat*. *Cat* contains several features, shown in (40), including the feature *HEAD*. In the Grammar Matrix core grammar, there are many *HEAD* values representing the various parts of speech (noun, verb, adjective, etc.), as well as all of the combinations of these parts of speech organized into a hierarchy of types (such as verb **and** adjective: +vj, etc.). *HEAD* is copied up a parse tree through the Head Feature Principle (Pollard and Sag 1994):

## (40) Definition of CAT

$$\left[ \begin{array}{c} \text{SYNSEM} \\ \text{LOCAL} \\ \text{CAT} \end{array} \left[ \begin{array}{c} \text{HEAD} \\ \text{VAL} \\ \text{POSTHD} \end{array} \left[ \begin{array}{c} \left[ \begin{array}{cc} \text{MOD} & \text{list} \\ \text{KEYS} & \text{keys\_min} \end{array} \right] \\ \left[ \begin{array}{cc} \text{SUBJ} & \text{list} \\ \text{SPR} & \text{list} \\ \text{COMPS} & \text{list} \end{array} \right] \\ \text{bool} \end{array} \right] \right] \right]$$

## (41) The Head Feature Principle

$$\left[ \begin{array}{c} \text{SYNSEM.LOCAL.CAT.HEAD } \boxed{1} \\ \text{HEAD-DTR.LOCAL.CAT.HEAD } \boxed{1} \end{array} \right]$$

*Cat* stores references to the subject, specifier, and complements of a type as separate lists in the type *val*. These references are used to connect interacting phenomena.

For instance, a lexical rule representing semantic feature agreement between a verb and its subject will constrain the semantic index of the first item on the subject list of the verb. The subj-head phrase structure rule identifies the first item on the subject list of the head (the verb, in this case) with the syntax and semantics (SYNSEM) feature structure of the non-head daughter. Because the semantic index is stored within the syntax and semantics feature structure, this identification also identifies the semantic index of the subject (the non-head daughter) with the semantic index of first item on verb's subject list. This identification only unifies if the features specified on the verb unify with the features on the semantic index of the subject. This properly constrains the application of the phrase structure rule to cases where the subject and the verb's specifications unify.

Other *cat* features are core to my analysis. POSTHD is a boolean feature for constraining the direction of modification, where types defined with [POSTHD +] are constrained to be after the head, and vice versa. MOD is a feature storing a list of modificands of the lexeme, such as a pointer to the noun from the adjective that modifies it. The MOD list and the valence lists that in practice stores values of *local* so that a constraint on the semantic index of a modificand might look like the following:

$$(42) \left[ \text{SYNSEM.LOCAL.CAT.HEAD.MOD} \left\langle \left[ \text{LOCAL.CONT.HOOK.INDEX.PNG 3rd} \right] \right\rangle \right]$$

### *The Lexical Integrity Hypothesis*

Bresnan and Mchombo 1995 introduce the Lexical Integrity Hypothesis that distinguishes the systems for word formation from the systems for phrase formation. In order terms, there is a strong division between morphology and syntax. This is encoded in the Grammar Matrix as a constraint on syntactic phrase structure rules that their daughters must be *infl-satisfied*, a collection of *luk*-valued features defined on a language-specific basis (Goodman 2013). Morphological rules (lexical rules) apply to a given stem, checking off required inflection for each required affix class. Once all the required affix classes have applied, the stem unifies with *infl-satisfied* and can be the daughter of syntactic rules. Importantly, Goodman 2013 posit that the definition of *infl-satisfied* can vary among word classes, allowing different part of speech types and different word types within a part of speech to vary on required inflection.

These constraints lead to a division of three separate morphological constructs: *position classes*, *lexical rule types*, and *lexical rule instances*. Less formally, position classes are morphological *slots* that take minimal sets of morphemes. For instance, a language might have position class on verbs for tense, where morphemes for past, present, and future tense cannot cooccur. Lexical rule types are collections of features for each position class, such as [TENSE *past*] on verbs or [GENDER *feminine*] on nouns. Lexical rules types are also supertypes for lexical rule instances, a set of affixes associated with a given lexical rule type and its features. Many times, lexical rule types only have one instance, and importantly, the orthographical form (i.e. the change to the stem) of lexical rule instances can be null.

Formally, in the Grammar Matrix, position classes are defined as a subtype of a basic lexical rule, specify the possible inputs to the rule (under DTR), identify unchanged inflection flags from the input to the output (from DTR to INFLECTED), and constrain changed inflection flags to +. This way, as required rules apply, the stem comes closer to unifying with *infl-satisfied*. Lexical rule types take the proper position class as their supertype, as well as *infl-lex-rule* if they have affixes or *const-lex-rule* if they do not, and constrain any features relevant for this lexical rule type. Lastly, lexical rule instances define an affix (if applicable), its position (prefix or suffix), and have a lexical rule type as their supertype.

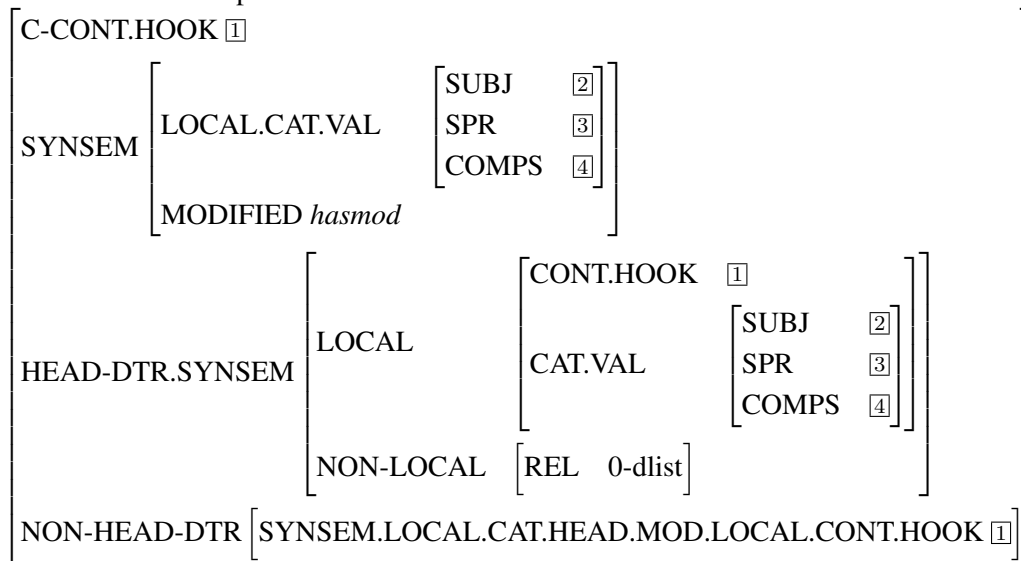
### *Special constraints*

In addition to type description discussed above, there are two additional types of lists that are important to the analysis of adjectives. First is the empty list, represented as  $\langle \rangle$ . Second is *cons*, which constrains a list to be non-empty, but does not constrain what values it might have. When a value in a list is constrained, this does not unify with  $\langle \rangle$ . This way, particular unifications can be blocked conveniently.

### *The Head Modifier Phrase*

Modifiers syntactically combine with the head of a phrase via the Head Modifier, or head-mod, phrase structure rule. In terms of adjectives, the adjective is the modifier and the head is a noun, with the resulting phrase a nominal constituent. In this way, the result of the Head Modifier phrase has the same distribution as a noun, able to combine with determiners, other adjectives, or other nominal constructions without additional rules. The Head Modifier phrase structure rule is *head compositional*, meaning that it inherits from the type *head-compositional*, which identifies the external semantic features (the hook, including the external argument and the semantic index) of the head daughter with the external semantic features of the phrase's semantics (the C-CONT, or construction content), shown in (43).

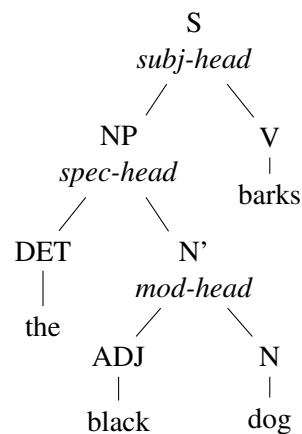
(43) The Head Modifier phrase



The Head Modifier phrase identifies the non-head daughter's modificand's external features (its *hook*, comprised of the external argument and semantic index) with the head daughter's modificand's external features and the phrase structure rule's external features. This also identifies the semantic index of these three: the non-head daughter's modificand's index, the head daughter's index, and the rule's index.

In English, adjectives combining via the Head Modifier phrase result in syntactic structures such as the following:

(44) Sample Head Modifier phrase syntactic structure



### *Raising and Copulas*

Copulas are analyzed as a raising verbs, following the analysis in Pollard and Sag 1994. Raising verbs are identified as those verbs which take a clausal complement and “pass up” the subject of their complement to their subject. Raising is codified through lexical entries for raising verbs that identify the external argument of the verb's complement with the semantic index of the verb's subject.

(45) Type definition for raising constraints

$$\left[ \begin{array}{c} \text{trans-first-arg-raising-lex-item} \\ \text{SYNSEM.LOCAL.CAT.VAL} \left[ \begin{array}{c} \text{SUBJ} \left\langle \text{LOCAL.CONT.HOOK.INDEX} \quad \boxed{1} \right\rangle \\ \text{COMPS} \left\langle \text{LOCAL.CONT.HOOK.XARG} \quad \boxed{1} \right\rangle \end{array} \right] \end{array} \right]$$

Copulas are analyzed as semantically empty raising verbs, such that they have no semantic contribution to the sentence, but do link their subject and complement. The copula also raises the semantics of its complement, identifying the semantic index and external argument of its complement with its own:

- (46) Type definition for semantically empty raising lexical types to identify their semantic index with their complement's semantic index

$$\left[ \begin{array}{l} \text{raise-sem-lex-item} \\ \text{SYNSEM.LOCAL} \left[ \begin{array}{l} \text{CONT} \left[ \begin{array}{l} \text{RELS } \langle ! ! \rangle \\ \text{HOOK } \boxed{1} \end{array} \right] \\ \text{CAT.VAL.COMPS } \left\langle \left[ \text{LOCAL.CONT.HOOK } \boxed{1} \right], \dots \right\rangle \end{array} \right] \end{array} \right]$$

### *The Head Subject Phrase*

The Head Subject phrase, or head-subj phrase, combines verbs with their subjects. This is relevant in two predicative adjective constructions: first is adjective as stative predicate, where the subject directly combines with the adjective through the Head Subject phrase; second is combining the subject with the copula.

### *Summary of HPSG*

DELPH-IN style grammars are typed feature structure grammars, where types are defined in a typed feature structure hierarchy. Feature structures of type *sign* have several key substructures, including a syntactic feature bundle (*cat*) and a semantic feature bundle (*cont*). *Sign* types come in three relevant roles: lexical types, lexical rules, and phrase structure rules. When types are combined, either through inheritance, the application of a lexical rule, or the application of a phrase structure rule, their constraints are checked to see if they unify with each other. A type combination is licensed if all of the constraints involved unify. A token is licensed if it can be mapped to a lexical type and if affixes, if any, can be analyzed by lexical rules and that these lexical rules unify with the lexical type. An utterance is licensed if each token unifies with a daughter of a phrase structure rule, and



these phrase structure rules combine (possibly via additional phrase structure rules) to result in a phrase structure rule that unifies with the root condition.

Grammatical description is done through type definitions, which utilizes multiple inheritance. Lexical types and lexical rules make reference to adjacent types through their valence lists, which, along with phrase structure rules, combine lexemes into phrases.

### 3.2.2 MRS

Minimal Recursion Semantics is a flat semantic representation made up of Elementary Predications (EPs, semantic relations). There are four core parts of an MRS: *hook*, *rels*, *hcons*, and *icons*.

1. *hook*: Core values of a sign externally visible to other signs
2. *rels*: List of EPs
3. *hcons*: Scope constraints on EPs
4. *icons*: Individual constraints for information structure

When viewed as a graph, MRS is comprised of EPs as nodes with shared arguments as edges.

An example of an MRS, for the English sentence *the black dog barks*, is as follows in (47).<sup>2</sup>

#### *HOOK and RELS*

The *hook* contains, among other attributes, the semantic index (INDEX) and external argument (XARG). The semantic index is the salient nominal instance or event associated with the constituent the MRS represents. The external argument is the external, controlled argument of a phrase, and is usually identified with either the subject or modificand.

RELS, or relations, is an unordered list of relations “collected” at unification. Lexical types such as adjectives have only one relation stored in the relations list, and these relations are appended by phrase structure rules up the parse tree. The bag of relations in the relations list is a key piece of the MRS representation. Each relation has a list of ARGn arguments.

---

<sup>2</sup>This MRS is generated by the English Resource Grammar version 1212; ICONS is not implemented in this version of the ERG

(47) MRS for *the black dog barks* from the ERG.

$$\left[ \begin{array}{l} \text{TOP } [h1] \\ \text{INDEX } [e3] [E.TENSE \text{ past}] \\ \text{RELS } \left\langle \begin{array}{l} \left[ \begin{array}{l} \text{"_the\_q\_rel"} \\ \text{LBL } [h4] \\ \text{ARG0 } [x6] \\ \text{RSTR } [h7] \\ \text{BODY } [h5] \end{array} \right], \left[ \begin{array}{l} \text{"_black\_a\_1\_rel"} \\ \text{LBL } [h8] \\ \text{ARG0 } [e9] \\ \text{ARG1 } [x6] \end{array} \right], \\ \left[ \begin{array}{l} \text{"_dog\_n\_1\_rel"} \\ \text{LBL } [h8] \\ \text{ARG0 } [x6] \left[ \text{PNG } \left[ \begin{array}{l} \text{PERS } 3 \\ \text{NUM } \text{sg} \end{array} \right] \right] \end{array} \right], \left[ \begin{array}{l} \text{"_bark\_v\_1\_rel"} \\ \text{LBL } [h2] \\ \text{ARG0 } [e3] \\ \text{ARG1 } [x6] \end{array} \right] \end{array} \right\rangle \\ \text{HCONS } \langle [h1] \text{ qeq } [h2], [h7] \text{ qeq } [h8] \rangle \end{array} \right]$$

In nouns, verbs, adjectives, and adverbs in the Grammar Matrix and the ERG, the ARG0 is identified with the semantic index of the lexeme, being the salient nominal instance or event (Flickinger 2000; Bender et al. 2002). Similarly, the ARG1 is usually identified with the external argument of a lexeme.

In the English sentence *the black dog barks*, and its MRS in (47), there are four items on the relations list, each representing a word in the English sentence to a relation in the MRS. The relation “\_big\_a\_1\_rel” has an ARG1 which is coindexed with the ARG0 of “\_dog\_n\_1\_rel”. This coindexation is due to the identification of the index of first item on the adjective’s modificand list (its MOD list) with the index of the head of head-mod phrase, in this case, the “\_dog\_n\_1\_rel” relation. This constraint on the head-mod phrase is display in (43).

## INDEX

The semantic index comes in two primary varieties. First is the **referential index**, a type used for noun-like references, defined as:

$$(48) \begin{bmatrix} ref-ind \\ PNG\ png \end{bmatrix}$$

where the type *png* is defined at a language-specific level, but generally contains features like person, number, gender, animacy, or a combined person-number hierarchy.

Second is the **event** index, which refers to events such as verbal or adjectival relations and is defined as:

$$(49) \begin{bmatrix} event \\ E\ tam \\ SF\ iforce \end{bmatrix}$$

where *iforce* is the illocutionary force of the phrase, such as proposition, question, or command, and *tam* is defined as:

$$(50) \begin{bmatrix} tam \\ TENSE\ tense \\ ASPECT\ aspect \\ MOOD\ mood \end{bmatrix}$$

This way, semantic features such as person, number, and gender are constrained to nominal “referential” types and semantic features such as tense, aspect, and mood are constrained to “event” types, and these features can be accessed for agreement, inflection, or other phenomena through the identification of the semantic index, usually with a valence position such as an item on the subject or complement list, or on the modificand.

### 3.2.3 Summary

This section discussed some important background in HPSG and MRS to adjectives in the Grammar Matrix. HPSG is a unification based typed feature structure built around a strongly lexicalist type hierarchy. Key values of the semantic analysis of adjectives are stored as a *hook* feature structure at a type’s core semantic content, or within a pointer to a type’s modificand or subject. MRS is a flat

semantic representation, a collection of relations connected via shared arguments. The ARG0 and ARG1 are key arguments to each relation, representing the core salient nominal instance or event and the external argument, respectively.

### 3.3 *Adjectives in HPSG Grammars*

The goal of the next section is to compare adjectives across several implementations in the DELPH-IN formalism. Three grammars, the English Resource Grammar (Flickinger 2000), Jacy (Japanese, Siegel and Bender 2002), and gCLIMB (German, Fokkens 2011) are compared and contrasted with the existing Grammar Matrix definitions. It is important to note that Jacy was built alongside the Grammar Matrix, and their history is intertwined, while the Grammar Matrix is built on top of the ERG and gCLIMB on top of the Grammar Matrix.

#### 3.3.1 *ERG*

The English Resource Grammar (ERG) is a broad coverage deep linguistic grammar of English, and is generally considered the largest DELPH-IN grammar. Subsequently, it is here first that I look to contrast the type definitions of the Grammar Matrix. It is important to note that the Grammar Matrix's core type definitions are based on the ERG (Bender et al. 2002).

The ERG is also the only large DELPH-IN grammar to implement attributive only and predicative only adjectives. Some examples of these phenomena follow:

- (51)
- a. regular
    - The **big** dog barked
    - The dog is **big**
  - b. attributive only
    - A **mere** beginner won
    - \*The winner is **mere**
  - c. predicative only
    - \*The **awake** dog barked
    - The dog is **awake**

Note in these examples, *mere* and *awake* cannot appear in both syntactic constructions. These are termed **attributive only** and **predicative only** adjectives, respectively. In order to capture these

syntactic differences, the ERG makes use of two syntactic features: PRD and POSTHD, discussed in §3.3.1 and §3.3.1 respectively.

However, there is also a third environment where adjectives appear consistently in English that is captured in the ERG. Consider the following:

- (52)
- a. the **dirty** dog barked
  - b. the house is **dirty**
  - c. everything **dirty** should be cleaned
  - d. \*the **ablaze** house roared
  - e. the house is **ablaze**
  - f. anything **ablaze** should be put out
  - g. the **mock** interview went well
  - h. \*the interview was **mock**
  - i. \*anything **mock** should be thrown out

The following sections will discuss these.

### *PRD*

In English, the class of adjectives that cannot appear in the standard attributive construction does appear in the post head constructions such as those *ablaze* above. The class of adjectives that cannot appear in the predicative construction, including *mock* above, also cannot appear in the post head construction. Finally, regular adjectives that can appear in the standard attributive and predicative construction can also appear in this post head construction.

In order to capture this, the ERG uses the value [PRD +] to distinguish adjectives that can be in the predicative and post head constructions, and [PRD –] to distinguish adjectives that cannot be. Adjectives left unspecified can appear in either. Therefore, the three classes of adjectives in English are distinguished.

### *POSTHD*

POSTHD, or post head, is defined as a boolean feature of *cat*:

- (53)  $\left[ \text{SYNSEM.LOCAL.CAT.POSTHD} \quad \text{bool} \right]$

POSTHD is used to constrain the direction of modification, such that [POSTHD +] modifiers can only unify with phrase types where the head is the first item in the phrase, and vice versa for [POSTHD –].

### *Adjectives in the ERG*

With the previous constraints on PRD and POSTHD, the syntactic constraints on adjectives in the ERG can be summarized as:

Regular (unconstrained)

$$(54) \left[ \text{SYNSEM.LOCAL.CAT} \begin{bmatrix} \text{HEAD.PRD} & \text{bool} \\ \text{POSTHD} & \text{bool} \end{bmatrix} \right]$$

Attributive only

$$(55) \left[ \text{SYNSEM.LOCAL.CAT} \begin{bmatrix} \text{HEAD.PRD} & - \\ \text{POSTHD} & - \end{bmatrix} \right]$$

Predicative only

$$(56) \left[ \text{SYNSEM.LOCAL.CAT} \begin{bmatrix} \text{HEAD.PRD} & + \\ \text{POSTHD} & + \end{bmatrix} \right]$$

While attributive and predicative adjectives happen to pattern in parallel in English, POSTHD and PRD are still reserved for their unique use cases, i.e. POSTHD constraining the direction of modification in modifiers and PRD constraining whether an item can be predicative or not.

These constraints are organized into two type hierarchies in the ERG. The first is a hierarchy of lexical types (types shown in (57)-(61), hierarchy in (61)) and the second is a hierarchy of *synsem* types (shown in (62)). For the most part, types in the lexical hierarchy constrain the value of their *synsem* value to one of the types in the *synsem* hierarchy.

(57) ERG basic adjective definition

$$\left[ \begin{array}{l} \text{norm-adj-word} \\ \text{SYNSEM.LOCAL.CAT.HEAD adj} \end{array} \right]$$

(58) ERG basic predicative only definition

$$\begin{bmatrix} \textit{pred\_only\_adj\_lex} \\ \text{SYNSEM intrans\_pred\_adj\_synsem} \end{bmatrix}$$

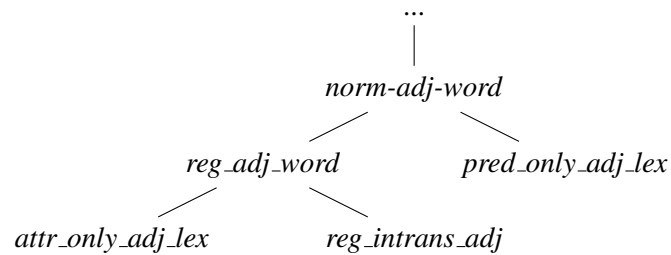
(59) ERG basic attributive adjective definition

$$\begin{bmatrix} \textit{reg\_adj\_word} \\ \text{SYNSEM} \begin{bmatrix} \text{LOCAL} \begin{bmatrix} \text{CAT.HEAD.MOD} \langle \text{LOCAL.CONT.HOOK.INDEX } \boxed{1} \rangle \\ \text{CONT.HOOK.INDEX } \boxed{1} \end{bmatrix} \\ \text{LKEYS.KEYREL.ARG1 } \boxed{1} \end{bmatrix} \end{bmatrix}$$

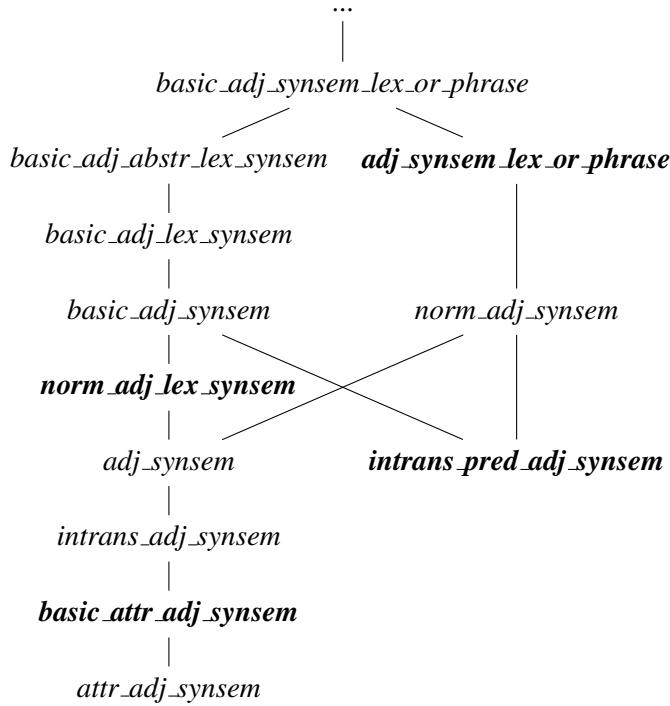
(60) ERG basic attributive only definition

$$\begin{bmatrix} \textit{attr\_only\_adj\_lex} \\ \text{SYNSEM } \textit{basic\_attr\_adj\_synsem} \end{bmatrix}$$

(61) ERG basic regular adjective

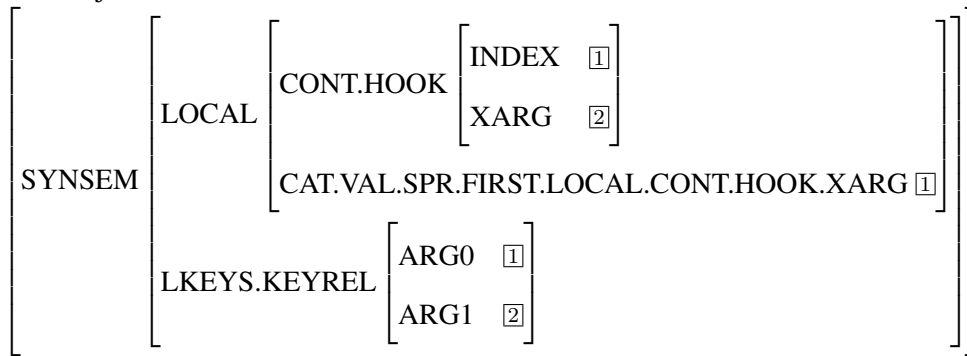
$$\begin{bmatrix} \textit{reg\_intrans\_adj} \\ \text{SYNSEM } \textit{intrans\_adj\_synsem} \end{bmatrix}$$


The *synsem* hierarchy has four key types: *basic\_attr\_adj\_synsem*, *norm\_adj\_lex\_synsem*, *intrans\_pred\_adj\_synsem*, and *adj\_synsem\_lex\_or\_phrase*. The primary hierarchy is below:

(62) ERG *synsem* types

The various types built into the adjective hierarchy in the ERG have a few commonalities. Primarily, they all share these same constraints:

## (63) Core adjectival constraints in the ERG



The identification of the type's specifier's external argument with the type's semantic index is for degree specifiers, such as *the **very** black dog barked*. These constraints serve as a bridge between the syntax and semantics, connecting the MRS' relations list (ARG0-ARGn) to their syntactic arguments: the semantic index to ARG0 and the syntactic external argument (typically the subject or modificand) with the semantic ARG1.

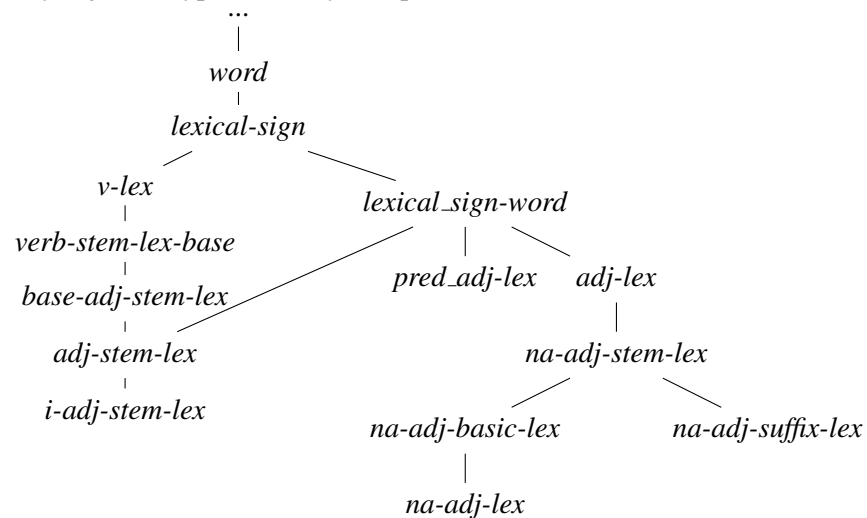


### 3.3.2 Jacy

Jacy is a large implemented grammar of Japanese (Siegel and Bender 2002). While Jacy predates the Grammar Matrix, it has been retrofitted onto an earlier version of the Grammar Matrix core grammar and much of its development has been in parallel to the Grammar Matrix, both in terms of syntax and semantics.

Japanese is of interest in that there are several distinct classes of adjectives, primarily those ending in *--i* and *--na* in their attributive form. These classes of adjectives are represented in the following type hierarchy:

(64) Jacy adjective type hierarchy (simplified)



Note that like the ERG, *pred\_adj-lex* is defined high in the tree, apart from other adjective types.

As mentioned previously, Japanese attributive adjectives are often analyzed as predicative adjectives in relative clause constructions (Nishiyama 1999; Siegel and Bender 2002). Consider the following examples:

## (65) a. Japanese predicative NP

sono hon wa shyousetsu desu  
this book TOP novel COP

‘This book is a novel’ [jpn] (Dalrymple et al. 2004)

## b. Japanese predicative adjective

sono hon wa akai desu  
this book TOP red COP

‘This book is red’ [jpn] (Dalrymple et al. 2004)

## c. Japanese relative clause as NP modifier

John wa Mary ga kaita hon-o yonda  
John TOP Mary NOM write-PST book ACC read

‘John read the book that Mary wrote’ [jpn] (Saito 1985)

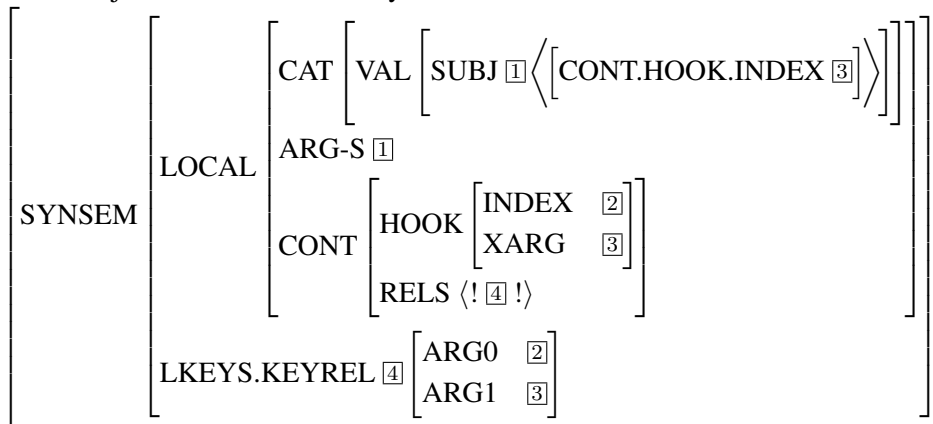
## d. Japanese relative clause/attributive adjective as NP modifier

Hanako ga san satsu takai hon o katta  
Hanako NOM three NUM-CL expensive book ACC buy.PST

‘Hanako bought three expensive books’ [jpn] (Ko 2005)

This is encoded as the external argument of an adjective being identified with the index of its subject (as opposed to its modificand):

## (66) Basic Adjective Definition in Jacy



Of note in Jacy is the identification of the semantic index (INDEX) with the ARG0 and the external argument (XARG) with the ARG1, consistent with the ERG. It is also of note that Jacy,

similarly to the ERG, defines *pred\_adj-lex* separately from the primary hierarchy. While this type is used for several lexemes in Jacy, lexemes defined as instances of this type are always homophonous with an instance of another adjectival type.

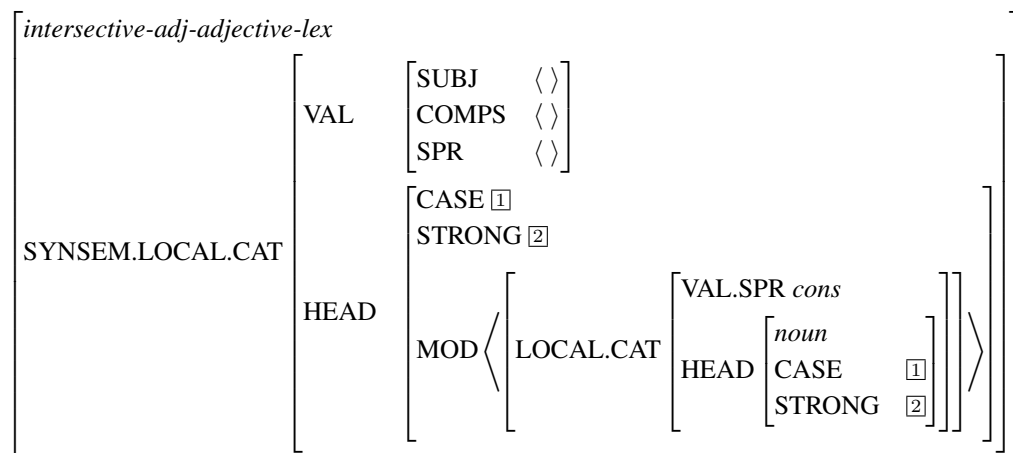
### 3.3.3 gCLIMB

CLIMB is a meta-grammar engineering environment, built on top of the Grammar Matrix, and designed to enable comparing competing analyses of phenomena (Fokkens 2011, to appear). Several grammars have been developed using CLIMB, including a family of grammars known as gCLIMB, which include grammars of German, Dutch, and Danish. The German grammar is the primary development target, and includes an implementation of adjectives on top of the Grammar Matrix.

German adjectives are illustrative of a category of adjectival behavior where the adjective agrees with its modificand in attributive constructions, but does not agree with its subject in predicative constructions.

First, the basic definition of agreement in gCLIMB can be seen in the following type:

(67) Basic adjective definition in gCLIMB with identification of semantic features.



This type constrains the case value of the modificand to be the same as the case value of the adjective. This way, only morphemes with the appropriate case value unify with the adjective, ensuring that only the proper agreement morphemes are legal. As discussed in §3.2.1, to ensure that any required agreement position classes are realized, the Grammar Matrix customization system

provides gCLIMB the type *infl-satisfied*, which is defined as a collection of *luk* flags, one for each obligatory position class. Phrase structure rules constrain their daughters to be *infl-satisfied*, which, in the case of agreement inflection, results in only adjectives with all of the obligatory affixes to be licensed in utterances.

This type also differs slightly from adjective types in the existing Matrix core grammar by constraining the modificand to be a noun with the constraint [MOD.HEAD *noun*]. The existing Matrix core types do not constrain the part of speech type of the modificand this way.

Second, gCLIMB provides an example of agreement only applying to the modificand of an adjective, and not to the subject. In order to model and enable this behavior, gCLIMB makes use of a basic lexical type and a position class which takes as input the basic lexical type. First is the basic adjective type, as shown above in (67), which is specified as a supertype to all adjectival lexical instances. These underspecified instances can be used as predicative adjectives along with the copula. Second is the position class and associated lexical rules that takes the basic adjective as its daughter. The lexical rule types are subtypes of the following type (slightly reconfigured and simplified):

(68) Adjective lexical rule supertype from gCLIMB

$$\left[ \begin{array}{l} \text{adj-inflection-lex-rule-super} \\ \text{INFLECTED.ADJ-INFLECTION-FLAG +} \\ \text{DTR } \textit{intersective-adj-adjective-lex} \end{array} \right]$$

Using the Grammar Matrix system of requiring morphemes with *infl-satisfied*, each adjective must have an instance of this lexical rule. The agreement inflection morphemes constrain their daughter to be [PRD −], while a null affixing predicative rule constrains its daughter to [PRD +]. The copula then constrains its complement to be [PRD +]. This way, agreement inflection is attributive only; agreement inflection is incompatible with the predicative construction.

- (69) Exemplary agreement rule from gCLIMB (Fokkens 2011, to appear):

$$\left[ \begin{array}{c} \text{strong-dat-masc-neut-sg-lex-rule} \\ \\ \text{SYNSEM.LOCAL.CAT.HEAD} \left[ \begin{array}{c} \text{PRD } - \\ \text{CASE } \textit{dat} \\ \text{STRONG } + \\ \text{MOD.FIRST...INDEX.PNG} \left[ \begin{array}{c} \text{NUM singular} \\ \text{GEND non-feminine} \end{array} \right] \end{array} \right] \end{array} \right]$$

- (70) Non-inflecting rule for predicative adjectives:

$$\left[ \begin{array}{c} \text{pred-lex-rule} \\ \text{SYNSEM.LOCAL.CAT.HEAD.PR.D } + \end{array} \right]$$

Thirdly, case marking is constrained separately from agreement features. As shown in (67), the adjective's case value is identified with its modificand's case value, which, when combined with the constraints in (69), results in the proper affixes being associated with the proper noun cases. Because this constraint is only between the modificand and the adjective, this case agreement does not apply to the subject, as desired.

Lastly, gCLIMB has two separate constraints to semantically connect the adjective with its argument, whether subject or modificand. First, to properly connect the semantics in predicative constructions, adjectives are constrained to identify their external argument (XARG) with their first semantic argument (ARG1), as codified in (71). Then, the copula, analyzed as a raising verb, identifies the adjective's external argument (the copula's complement's external argument) with the semantic index of the copula's subject, as codified in (72).

- (71) gCLIMB adjective identification between semantic external argument and first semantic argument:

$$\left[ \begin{array}{c} \text{int-mod-adj-lex} \\ \text{SYNSEM} \left[ \begin{array}{c} \text{LOCAL.CONT.HOOK.XARG } \boxed{1} \\ \text{LKEYS.KEYREL.ARG1 } \boxed{1} \end{array} \right] \end{array} \right]$$

(72) gCLIMB raising supertype for copula definition:

$$\left[ \text{SYNSEM.LOCAL.CAT.VAL} \left[ \begin{array}{l} \text{SUBJ} \quad \langle \text{LOCAL.CONT.HOOK.INDEX} \quad \boxed{1} \rangle \\ \text{COMPS} \quad \langle \text{LOCAL.CONT.HOOK.XARG} \quad \boxed{1} \rangle \end{array} \right] \right]$$

Second, gCLIMB adjectives inherit from the type *intersective-mod-lex*, which identifies the semantic index of the adjective's modificand with the first semantic argument of the adjective. This way, both the adjective's subject's semantic index and the adjective's modificand's semantic index are identified with the first semantic argument of the adjective (the ARG1):

$$(73) \quad \left[ \begin{array}{l} \text{intersective-mod-lex} \\ \text{SYNSEM} \left[ \begin{array}{l} \text{LOCAL.CAT.HEAD.MOD} \quad \left\langle \left[ \text{LOCAL} \left[ \begin{array}{l} \text{intersective-mod} \\ \text{CONT.HOOK.INDEX} \quad \boxed{1} \end{array} \right] \right] \right\rangle \\ \text{LKEYS.KEYREL.ARG1} \quad \boxed{1} \end{array} \right] \end{array} \right]$$

### 3.3.4 The Grammar Matrix & Language CoLLAGE

The Grammar Matrix has had adjectives implemented in the core grammar file for grammarians to extend since near its origin (Bender et al. 2002). Since then, over one hundred grammars have been built on top of the core grammar file, most of them produced in the grammar engineering course at the University of Washington (see Bender 2007 for more information). Bender 2014 introduces a subset of these grammars as Language CoLLAGE. Implementations of adjectives as engineered during the coursework are provided in Language CoLLAGE, acting as examples of predicative and attributive adjectives on top of the Grammar Matrix in several languages.

The core types used in the CoLLAGE grammars are in (74)-(77), which are organized into a type hierarchy in (78).

$$(74) \quad \left[ \begin{array}{l} \text{Grammar Matrix basic adjective type} \\ \text{basic-adjective-lex} \\ \text{SYNSEM.LOCAL.CAT.HEAD} \quad \text{adj} \end{array} \right]$$

- (75) Grammar Matrix type definition linking modificand to external argument

$$\left[ \begin{array}{c} \text{attrib-or-pred-lex} \\ \text{SYNSEM.LOCAL} \left[ \begin{array}{c} \text{CAT.HEAD.MOD} \langle \text{LOCAL...HOOK.INDEX } \boxed{1} \rangle \\ \text{CONT.HOOK.XARG } \boxed{1} \end{array} \right] \end{array} \right]$$

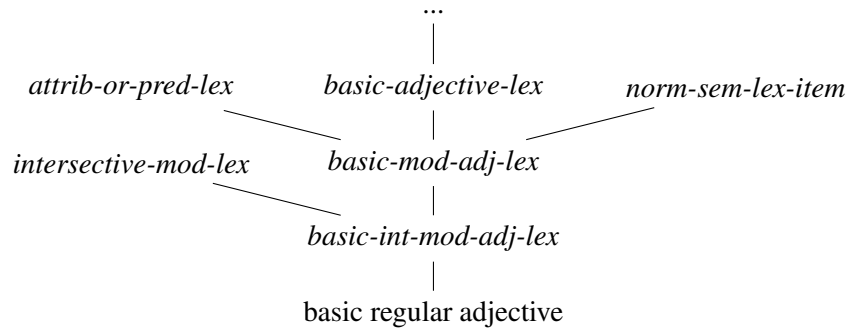
- (76) Grammar Matrix type to link semantic index with ARG0

$$\left[ \begin{array}{c} \text{norm-sem-lex-item} \\ \text{SYNSEM} \left[ \begin{array}{c} \text{LOCAL.CONT.HOOK.INDEX } \boxed{1} \\ \text{LKEYS.KEYREL.ARG0 } \boxed{1} \end{array} \right] \end{array} \right]$$

- (77) Grammar Matrix type to link modificand's index with ARG1

$$\left[ \begin{array}{c} \text{intersective-mod-lex} \\ \text{SYNSEM} \left[ \begin{array}{c} \text{LOCAL.CAT.HEAD.MOD} \langle \text{LOCAL...INDEX } \boxed{1} \rangle \\ \text{LKEYS.KEYREL.ARG1 } \boxed{1} \end{array} \right] \end{array} \right]$$

- (78) Grammar Matrix adjectival type hierarchy



This hierarchy is similar to the ERG's (see §3.3.1), but it does not support attributive only or predicative only adjectives. All adjectives inherit from *basic-int-mod-adj-lex*, which identifies the semantic index (INDEX) of the modificand (MOD) and external argument (XARG) with the ARG1, and identifies the semantic index of the adjective with ARG0. This links the syntax and semantics, linking the syntactic subject or modificand with the semantic argument, ARG1, and the event instance with the ARG0. This formulation enables the syntax to properly connect the MRS semantics.

Additionally, the Grammar Matrix customization system defines agreement through specifying the semantic features on the modificand or external argument of an adjective (shown in 79), or the semantic index of the subject of a copula (shown in 80):

$$(79) \left[ \begin{array}{l} 3pl\text{-}adjective\text{-}lex \\ SYNSEM...MOD...HOOK.INDEX.PNG \left[ \begin{array}{l} PER\ 3rd \\ NUM\ pl \end{array} \right] \end{array} \right]$$

#### Grammar Matrix Adjective and Copula Agreement

$$(80) \left[ \begin{array}{l} 3pl\text{-}copula\text{-}lex \\ SYNSEM...VAL.SUBJ...HOOK.INDEX.PNG \left[ \begin{array}{l} PER\ 3rd \\ NUM\ pl \end{array} \right] \end{array} \right]$$

The identification of the subject's semantic index or the modificand's semantic index with semantic index of the adjective undergoing the lexical rule then constrains agreement properly.

Lastly, CoLLAGE grammars also make use of the boolean feature POSTHD for controlling direction of modification, as well as the boolean feature PRD for specifying whether a lexical type can appear as a copula complement (similarly to the ERG, see §3.3.1 for further discussion).

### 3.4 Summary

There are several key takeaways from this survey of existing implementations of adjectives.

1. Identification of the first semantic argument (ARG1), external argument (XARG), and the modificand's semantic index (MOD..INDEX)
2. PRD boolean feature to constrain instances to be a copula complement or not
3. POSTHD boolean feature to constrain modification direction
4. Predicative adjectives are often the supertype of attributive adjectives (or a subtype of an attributive adjective supertype)
5. Adjective agreement can be achieved through constraining the proper values on the proper argument
6. Modificand specific agreement in German is achieved through use of bifurcated lexical rules, where inflected adjectives are constrained to be attributive and uninflected adjectives are constrained to be predicative
7. Case marking must be handled separately from semantic feature agreement



These key points are essential to the analysis as presented. The following chapter describes my analysis, which extends the existing implementations of adjectives discussed in this chapter to cover phenomena covered in Chapter 2.

## Chapter 4

### ANALYSIS

The existing Grammar Matrix core definitions include types for attributive and predicative adjectives, but are impoverished in some key areas:

1. There are no types to constrain adjectives from certain constructions in order to define attributive only and predicative only adjectives.
2. There is no distinction between copula complement predicative adjectives and stative predicate adjectives, or analysis of predicative adjectives more generally.
3. There is no analysis of copulas beyond the raising type definition.
4. The Grammar Matrix does not currently include predefined support for switching type languages.
5. There is no analysis of incorporated adjective stems.
6. There is no analysis of languages with construction-constrained argument agreement.

I organize these items in four separate groups. The first two groups are straightforward: the first group is the core adjectival phenomena, those relevant to most languages; the second group is the analysis of copulas.

The second two groups are given additional discussion due to their complexity and importance to my analysis. These two groups were referred to as adjectives in **switching languages** in Chapter 2, but I will now differentiate between them more strongly. The third group is mixed type languages, those languages with switching adjectives or multiple adjective types. The fourth group is construction-constrained adjectives, those adjectives which differ in their morphology depending on their syntactic behavior.

These four groups provide the structure for this chapter. §4.1 lays out the analysis of adjectives to be included in the core grammar and types to be distributed to language-specific grammars. §4.2 presents the analysis of copulas to be included in copula complement grammars. §4.3 covers the analysis of mixed type adjectives, including optional inflection and optionally copula complement adjectives. §4.5 details the analysis of adjective incorporation.

#### 4.1 Basic Adjectives in the Grammar Matrix

The initial inspiration for my analysis is the English Resource Grammar (ERG, Flickinger 2000) as it is the only large DELPH-IN grammar to analyze predicative only and attributive only adjectives. The ERG adjective hierarchy is complex, including many special constructions, such as comparatives, superlatives, etc. For the purposes of the Grammar Matrix, the constraints for the core types can be used as a basis for development. These types can be seen in (57)-(61).

The type hierarchy in (61), however, has many repeated constraints. This is partially because of the additional types not shown here and partially due to the evolution of the grammar. This type hierarchy can be rearranged such that repeated types on siblings are moved into supertypes, such as in (81)-(84), organized into the hierarchy in (85).

- (81) Simplified ERG Basic adjective definition

$$\left[ \begin{array}{c} \text{basic-adj-lex} \\ \text{SYNSEM} \left[ \begin{array}{c} \text{LOCAL} \left[ \begin{array}{c} \text{CONT} \left[ \begin{array}{c} \text{HOOK} \left[ \begin{array}{c} \text{INDEX } [1] \\ \text{XARG } [2] \end{array} \right] \\ \text{VAL} \left[ \begin{array}{c} \text{SPR.FIRST.LOCAL.CONT.HOOK.XARG } [1] \\ \text{COMPS } \langle \rangle \end{array} \right] \\ \text{HEAD } \textit{adj} \end{array} \right] \\ \text{LKEYS.KEYREL} \left[ \begin{array}{c} \text{ARG0 } [1] \\ \text{ARG1 } [2] \end{array} \right] \end{array} \right] \end{array} \right] \end{array} \right]$$

- (82) Simplified ERG attributive adjective definition

$$\left[ \begin{array}{c} \text{attr-adj-lex} \\ \text{SYNSEM} \left[ \begin{array}{c} \text{LOCAL.CAT.HEAD.MOD} \left\langle \left[ \text{LOCAL...INDEX } [2] \right] \right\rangle \\ \text{LKEYS.KEYREL.ARG1 } [2] \end{array} \right] \end{array} \right]$$

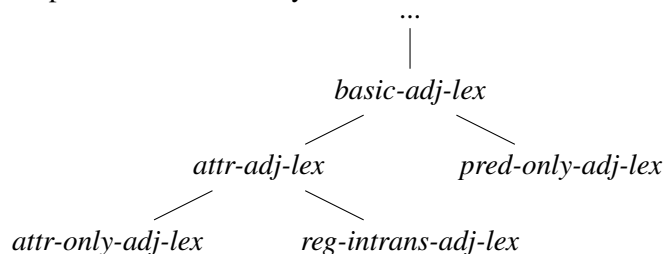
- (83) Simplified ERG predicative only adjective definition

$$\left[ \begin{array}{c} \text{pred-only-adj-lex} \\ \text{SYNSEM.LOCAL.CAT.HEAD.PRD } + \end{array} \right]$$

- (84) Simplified ERG attributive only adjective definition

$$\left[ \begin{array}{c} \text{attr-only-adj-lex} \\ \text{SYNSEM.LOCAL.CAT.HEAD.PRD } - \end{array} \right]$$

## (85) Simplified ERG hierarchy



However, as discussed in Chapter 2, there are several additional phenomena this hierarchy needs to account for. I break these into two groups: those relevant to a cross-linguistic analysis of adjectives broadly and those relevant to language-specific analyses:

## (86) Cross-linguistic phenomena:

1. Stative predicates
2. Scopal *vs.* intersective adjectives

## (87) Language-specific phenomena:

1. Copulas
2. Word order
3. Agreement and inflection
4. Mixed type adjectives
5. Construction constrained argument agreement
6. Incorporation

§4.1.1 discusses these cross-linguistic phenomena while §4.1.2 discusses these language-specific phenomena.

#### 4.1.1 A Cross-linguistic Analysis of Adjectives

The primary distinction not present in the Grammar Matrix is the distinction between predicative only and attributive only adjectives. While some languages have adjectives that participate in both sorts of constructions, others do not. Similarly, languages like English have adjectives that can only appear in one construction, such as *mock* in the attributive construction and *awake* in the predicative. The hierarchy based on the ERG presented above includes types for construction-constrained adjectives.

The first type currently not provided by the Grammar Matrix is for stative predicate adjectives. Stative predicates combine directly with their subject, so unlike attributive and copula complement adjectives, which forbid subjects, stative predicates require subjects. Furthermore, the external argument of a stative predicate adjective is its subject. Lastly, like other adjectives, the external argument is constrained to be a noun whose specifier has already been satisfied. These constraints are formalized in the following type:

$$(88) \left[ \begin{array}{l} \textit{stative-pred-adj-lex} \\ \\ \text{SYNSEM.LOCAL} \left[ \begin{array}{l} \text{CAT.VAL.SUBJ} \left\langle \begin{array}{l} \text{LOCAL} \left[ \begin{array}{l} \text{CAT} \left[ \begin{array}{l} \text{VAL} \left[ \begin{array}{l} \text{SPR} \quad \langle \rangle \\ \text{COMPS} \quad \langle \rangle \end{array} \right] \\ \text{HEAD} \textit{noun} \end{array} \right] \\ \text{CONT.HOOK.INDEX } \boxed{1} \end{array} \right] \right] \right] \end{array} \right\rangle \\ \text{CONT.HOOK.XARG } \boxed{1} \end{array} \right] \end{array} \right]$$

In order to include this type in the type hierarchy, it is necessary to formalize the distinguishing constraints between copula complement adjectives, stative predicate adjectives, and attributive adjectives. As discussed previously, the value PRD on HEAD is used in the ERG to constrain lexemes from combining with a copula. I continue to use this feature in my analysis, but in order to account for mixed type languages with copula complement, stative predicate, and attributive adjectives, additional constraints are necessary. I utilize [PRD –] to keep adjectives out of copula constructions, [SUBJ < >] to keep adjectives out of the head-subj phrase, and similarly [MOD < >] to keep adjectives out of the head-mod phrase. These are detailed below:

(89) Type definition for copula complement only adjectives:

$$\left[ \begin{array}{l} \textit{cop-comp-adj-lex} \\ \\ \text{SYNSEM.LOCAL.CAT} \left[ \begin{array}{l} \text{HEAD} \left[ \begin{array}{l} \text{MOD} \quad \langle \rangle \\ \text{PRD} \quad + \end{array} \right] \\ \text{VAL.SUBJ} \langle \rangle \end{array} \right] \end{array} \right]$$

(90) Type definition for stative predicate only adjectives:

$$\left[ \begin{array}{l} \textit{stative-pred-adj-lex} \\ \text{SYNSEM.LOCAL.CAT.HEAD} \left[ \begin{array}{ll} \text{MOD} & \langle \rangle \\ \text{PRD} & - \end{array} \right] \end{array} \right]$$

(91) Type definition for attributive-only adjectives:

$$\left[ \begin{array}{l} \textit{attributive-only-adj-lex} \\ \text{SYNSEM.LOCAL.CAT} \left[ \begin{array}{ll} \text{VAL.SUBJ} & \langle \rangle \\ \text{HEAD.PRD} & - \end{array} \right] \end{array} \right]$$

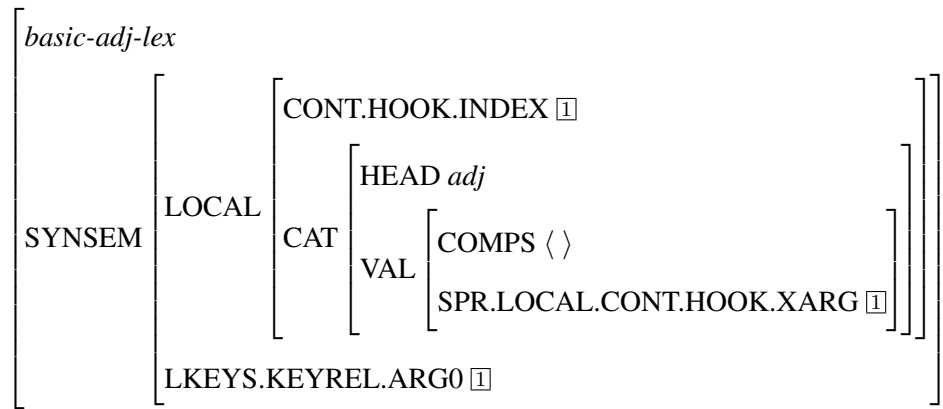
Note that the head-mod phrase requires a modificand value for the head daughter, which does not unify with [MOD  $\langle \rangle$ ]. Similarly, the head-subj phrase requires a subject value of the for the head daughter.

While the basic attributive definition, shown in (82), is not universally useful to languages, it seems to be universal enough to include in the core grammar. However, with the world's languages split more or less equally between stative predicate adjectives and copula complement adjectives (Stassen 2013), it seems appropriate to include these types in the language-specific type definitions.

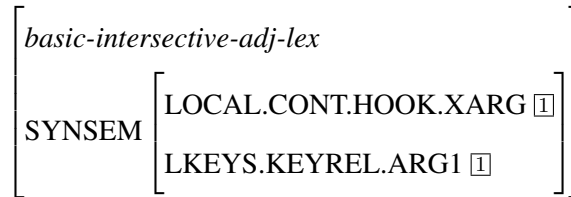
Lastly, while scopal adjectives are not considered in depth here, there is a basic type for them in the Grammar Matrix core grammar. The primary shared constraints between scopal and intersective adjectives are those described in the type *basic-adj-lex* in (81) above. The key difference between intersective and scopal adjectives is their semantic argument: intersective adjectives identify their ARG1 with their external argument (XARG), while scopal adjectives must account for scope constraints on their ARG1. Therefore, this constraint is separated out of the core adjective definition.

The resulting types are shown in (92) and (93), reorganizing the type hierarchy into (94).

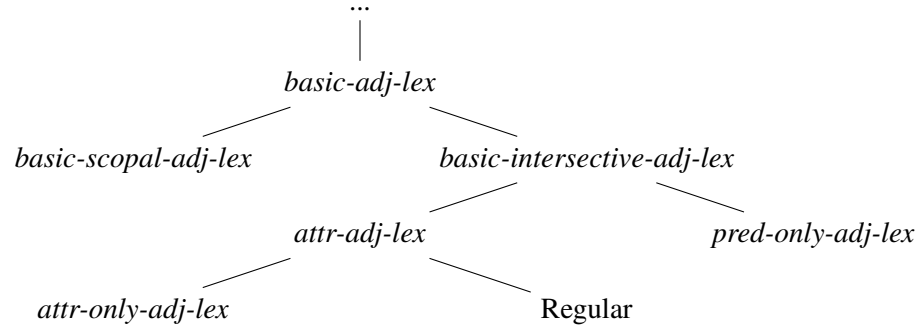
(92) Simplified basic adjective definition (compare with (85))



(93) Basic intersective adjective definition



(94) Reorganized type hierarchy (reorganized from (85))



While this analysis is sufficient for many languages tested, it fails to capture the linguistic facts of construction-constrained agreement inflection in languages such as German. In German, the adjective only agrees with its modificand in attributive constructions, and is only licensed without inflection in predicative constructions. The role of lexical and morphological constraints at the interface of syntax and morphology is an important theoretical question in designing an analysis of adjectives. Specifically, what is the best method to analyze an adjectival stem that is able to be attributive and predicative, but relies on specific morphology to license its syntactic behavior?<sup>1</sup>

<sup>1</sup>Of course, one could look at this from the other way around: adjectives are underspecified for syntactic behavior, and their morphology specifies their syntactic behavior, but the core issue still lies at the interface of syntax and morphology.

The issue in terms of HPSG is that because the lexical *regular* type in (85) is constrained to identify its modificand's index with its semantic argument, lexical rules cannot subsequently constrain their input to be predicative only ([MOD ⟨ ⟩]). Essentially, it is confusing to define an adjective in the lexicon as attributive and predicative, and then in the morphology constrain it to be predicative only. Instead, it seems to make more sense to keep these sorts of definitions in one place (in the morphology or lexicon), and provide a consistent mechanism to the user.

While so far I have relied exclusively on definitions in the lexicon to define the syntactic behavior of adjectives (attributive, predicative, etc.), at this point it appears that this restriction is no longer feasible. Rather, I propose that a cross-linguistic analysis of adjectives must rely both on definitions from the lexicon and additional syntactic constraints in the morphology.

Additionally, some languages have historically been characterized as having *attributive markers* or *predicative markers*. For instance, Kang 2005 analyzes Korean [kor] as having a predicative marker while Li 2008 analyzes Mandarin [cmn] as having an attributive marker. I consider these sort of phenomena identical to the sort of argument-constrained agreement in German, without the agreement.

It is not a novel idea to have morphological processes define the syntactic behavior of adjectives. gCLIMB (§3.3.3, Fokkens 2011) relies on morphological rules to define the syntactic behavior of adjectives. However, to my knowledge, gCLIMB and other analyses which rely on morphological constraints do so without reference to the theoretical implications involved. It seems noteworthy that adjective syntactic behavior is often defined morphologically as opposed to in the lexicon.

To implement this proposal, I first look to gCLIMB's analysis for German. gCLIMB's solution to this problem is to move constraints from lexical types (i.e. items in the lexicon) to constraints on lexical rules (i.e. morphological processes). However, this solution applied to adjectives in all languages would be cumbersome from an end-user engineering perspective. For languages that do not require this flexibility (most of those I investigated), the addition of lexical rules between the lexical types and their realization is simply unnecessary machinery.

This is where the machinery of the Grammar Matrix customization system is useful. Users can describe their language without consideration of these mechanisms, being guided by the online questionnaire, and the back-end system is able to build the required types dynamically. Specifically, the proper attributive, attributive only, and predicative only definitions are dynamically generated by

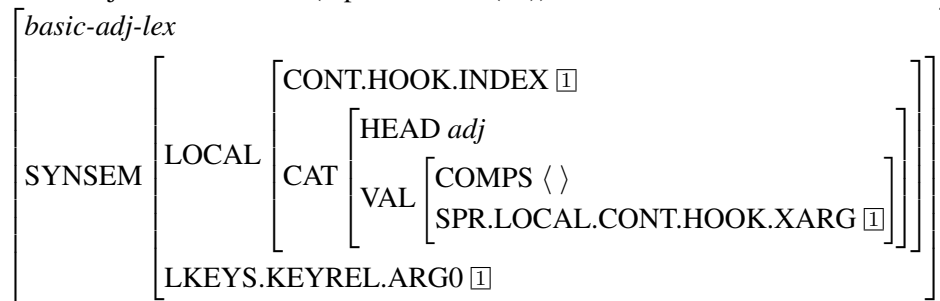


the customization system into either lexical types or lexical rules (or both, if applicable) and added to the language-specific type definitions. This results in the following situation in (102), where the types on the top appear in the Grammar Matrix core definition file, and the types on the bottom are automatically generated and put into the language-specific file.

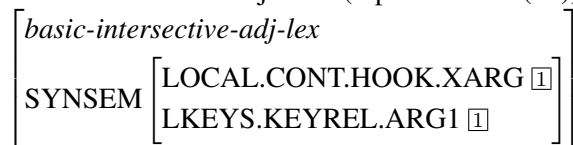
My final analysis is presented in the types in (95)-(101), represented in a less formal schematic in (102).<sup>2</sup>

### Core types:

- (95) Basic adjective definition (repeated from (92))

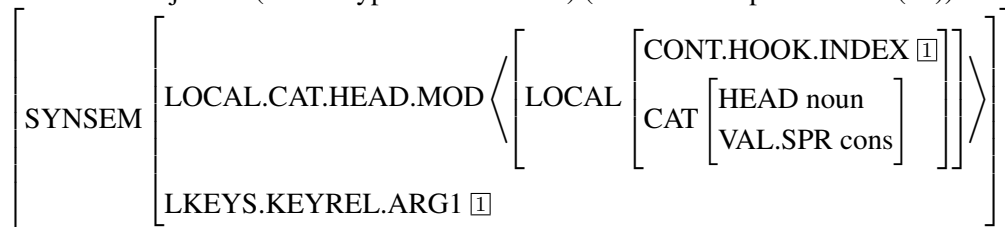


- (96) Basic Intersective Adjective (repeated from (93))



### Language-specific types:

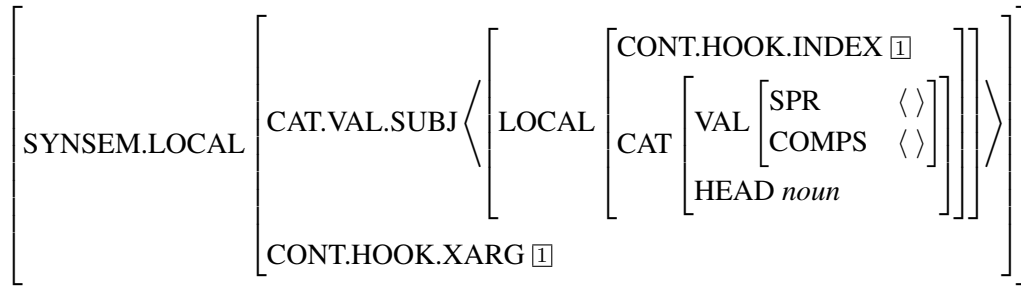
- (97) Attributive adjective (lexical type or lexical rule) (constraints repeated from (82))



- (98) Stative predicate adjective (lexical type or lexical rule) (constraints repeated from (88))

---

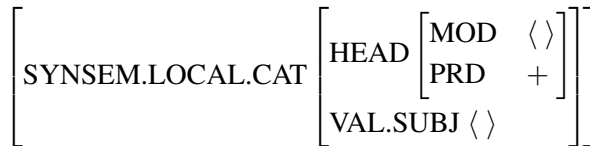
<sup>2</sup>This diagram is a representation of types as opposed to a strict definition of types in order to allow for the distinction between lexical rules and lexical types.



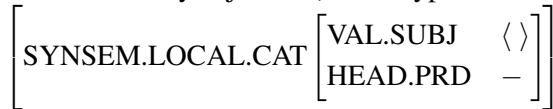
(99) Stative predicate only adjective (lexical type or lexical rule) (constraints repeated from (90))



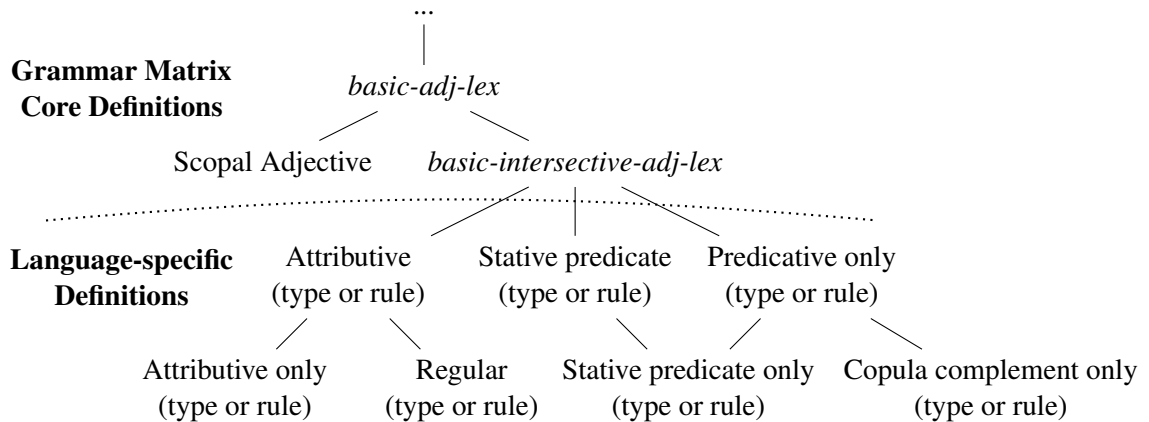
(100) Copula complement only adjective (lexical type or lexical rule) (constraints repeated from (89))



(101) Attributive only adjective (lexical type or lexical rule) (constraints repeated from (91))



(102) Representation of final core analysis:



In some languages, where appropriate, the stative predicate lexical type or lexical rule is also a supertype to the regular adjective lexical type or lexical rule.

Lexical rule types as part of the language-specific definitions are given the basic lexical rule supertype *add-only-no-ccont-rule*, while lexical types are given the supertype *basic-intersective-*

*adj-lex* as shown in (102). The language-specific lexical rule types are organized into a hierarchy as shown and are added as a supertype to the proper user-defined lexical rule types depending upon user choice (see §5.1.3 for specifics).

In summary, to accurately capture the range of variation of adjectives in the world's languages in Chapter 2, it is necessary to constrain syntactic behavior through morphological processes as opposed to definitions in the lexicon. I propose to use the functionality of the Grammar Matrix customization system capture this flexibility. Core types are stored in the static Grammar Matrix core grammar, while more specific definitions are generated by the customization system as either lexical types or lexical rules based on user input and included in the language-specific grammar file. This allows enough flexibility to capture the generalizations of the target phenomena while minimizing mechanisms.

#### *4.1.2 Language-Specific Constraints*

Along with the core definitions that are relevant in most languages as discussed above, there are variations in adjectives between languages and between types within a language. While the previous section, §4.1.1, discussed types thought to be relevant to most languages, mainly a basic adjective definition, basic attributive and predicative definitions, and types for keeping items out of particular syntactic constructions, this section is focused on phenomena known to be specific to individual languages and those that vary from language to language. Data for these phenomena are discussed more thoroughly above in Chapter 2.

This section will cover four core variations among languages:

1. Agreement and inflection
2. Attributive word order
3. Attributive adjectives being unique modifiers
4. Predicative adjective behavior (copula complement vs stative predicate)

#### *Agreement and Inflection*

Many languages have adjectives that agree with their external argument, as discussed in Chapter 2. I build upon the existing model and analysis of morphology in the Grammar Matrix, as discussed

in §3.2.1, from O’Hara 2008 and Goodman 2013. To summarize, this model separates morphology into three levels: position classes, lexical rule types, and lexical rule instances. Position classes are “slots” on a stem with a set of affixes in minimal distribution. Lexical rule types are types of affixes which fit into a particular position class, often also being defined with some constraints on semantic features. Lexical rule instances are the actual string: either a spelling-changing affix, known as inflectional rules, or **i-rule**; or an empty string, known as a lexical rule, or **l-rule**.

Adjectives and copulas are often marked for agreement with a referential feature, such as person, number, or gender agreement. Within the existing system of morphology, adjectives can be specified for agreement by constraining the external argument (XARG) of the adjective for the appropriate feature. These constraints are placed on lexical rule types. For instance, the following is an example of a lexical rule type constraining an adjective’s external argument’s gender to masculine:

(103) Example of agreement feature:

$$\left[ \begin{array}{l} \textit{masculine-lex-rule} \\ \text{SYNSEM.LOCAL.CONT.HOOK.XARG.PNG.GEND masculine} \end{array} \right]$$

Because the external argument (XARG) is identified with the ARG1 of the adjective, which in turn is identified with the modificand (MOD) and subject (SUBJ) in appropriate types, constraining the PNG features on the external argument enforces agreement between the adjective and the subject or modificand. This is done for all agreement features. For languages that have adjectives that agree with the subject but not the modificand or vice versa, additional specifications are required. See §4.4 for this analysis. For further discussion on the specific implementation of agreement constraints, see §5.1.2 and §5.2.2.

Similarly, adjectives and copulas often inflect for event features, such as tense, aspect, or mood. Event type features can be specified on lexical rules by constraining the semantic index of the adjective. The following is an example of a lexical rule type constraining the tense of an adjective to present:

(104) Example of event type features:

$$\left[ \begin{array}{l} \textit{adj\_tense-lex-rule-super} \\ \text{SYNSEM.LOCAL.CONT.HOOK.INDEX.E.TENSE present} \end{array} \right]$$

### *Attributive word order*

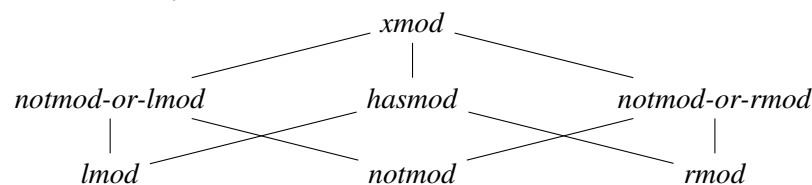
The Grammar Matrix core definitions make use of the POSTHD feature on HEAD along with ordered head-mod and mod-head phrase structure rules to specify word order on modifier phrases (see §3.2.1 for more information on the head-mod phrase). I continue using this infrastructure to constrain word order of attributive adjectives: the head-mod phrase structure rule constrains the modifier to be [POSTHD +] while the mod-head phrase structure rule constrains the modifier to be [POSTHD –].

These constraints on the phrase structure rules work together with constraints on individual adjective types to constrain the word order possibilities for attributive adjectives. Attributive adjectives that can only appear after their modificand are specified [POSTHD +], while adjectives that can only appear before their modificand are specified [POSTHD –]. Attributive adjectives that can be either pre-head or post-head do not specify a POSTHD value.

### *Unique Modification*

Some languages, such as Maori, only license one adjective per noun (see §2.3.4). The Grammar Matrix core definitions make use of a hierarchy of types to keep track of and constrain modifier types. A slightly simplified version of the hierarchy is given below:

(105) *xmod* hierarchy:



The mother of the head-mod phrase is constrained to be [MODIFIED *hasmod*], as shown in (43). Therefore, to constrain adjectives to only apply to modificands that have not already been modified, the following constraint is placed on the adjective type in the language-specific definitions:

(106) Unique modification constraint:

$$\left[ \text{SYNSEM.LOCAL.CAT.HEAD.MOD} \left\langle \left[ \text{MODIFIED } \textit{notmod} \right] \right\rangle \right]$$

### *Predicative Adjective Type*

As described above, users can define predicative adjectives as either copula complements, stative predicates, or optionally copula complement. Optionally copula complement adjectives are discussed in more detail in §4.3.2. Stative predicate types receive the stative predicate definitions either as a supertype or through a lexical rule as detailed in the previous section, shown in (88). Copula complement types are marked [PRD +].

Stative predicate adjectives can stand alone without a verbal predicate. Subsequently, stative predicate adjectives are often the primary predicate (specifically, the global INDEX) in the given sentence. Because unification relies on a root condition, languages with stative predicate adjectives must have the root condition of unification amended to include adjectival heads (see §3.2.1 for more on the root condition), a relaxation of the root condition. Grammars without stative predicates do not require this relaxation.

### *Summary of Language-specific Adjective Constraints*

This section has provided an overview of the analysis of language-specific constraints, including adjective agreement and inflection, attributive word order, unique modification, and predicative adjective behavior. For each of these, I rely on existing analyses included in the Grammar Matrix core grammar. Agreement and inflection relies on the morphology and morphotactics library. Attributive word order relies on the POSTHD feature built into the head-mod and mod-head phrase structure rules. Unique modification relies on the *xmod* hierarchy and the MODIFIED feature built into the Head Modifier Phrase supertype and the definition of SYNSEM. Predicative adjective behavior relies on the PRD feature, the subject list (SUBJ), the pointer to the modificand (MOD), and the definition of an empty list.

#### *4.1.3 Summary of Adjective Analysis*

I have presented my core analysis for including adjectives in the LinGO Grammar Matrix. Core to my analysis is the proposition that the syntactic behavior of adjectives is often defined during morphological processes instead of in the lexicon. This distinction is an interesting observation for languages which seem to have attributive or predicative markers, morphology on an adjective that

seems to mark its syntactic behavior. To model this distinction, I make use of the machinery of the Grammar Matrix customization system, keeping only the most basic adjective type definitions in the core grammar, while providing users with specific adjective definitions based on their input. Some of these definitions are in the form of lexical types while others take the form of morphological rules (lexical rules).

Additionally, in developing analyses for more language-specific behavior, I am able to rely on the foundation of the Grammar Matrix core grammar to capture the linguistic generalizations of many languages. I have presented an analysis of adjective morphology, word order, unique modification, and predicative behavior. I am able to capture these various phenomena using existing existing features and mechanisms of the Grammar Matrix core grammar.

The next section will provide an overview of my analysis of copulas.

## 4.2 *Copulas in the Grammar Matrix*

This section details the analysis of copulas in the Grammar Matrix. §4.2.1 covers the core definition of the copula, §4.2.2 covers the inflection and agreement analyses of copulas, and §4.2.3 covers an basic analysis of complement selection.

### 4.2.1 *Basic Copula Definition*

As covered in §2.4.4 and §3.3.3, the copula is a semantically empty raising verb that semantically connects its complement with its subject. This is codified in HPSG by identifying the semantic index of the subject with the external argument of the complement, shown in (107):

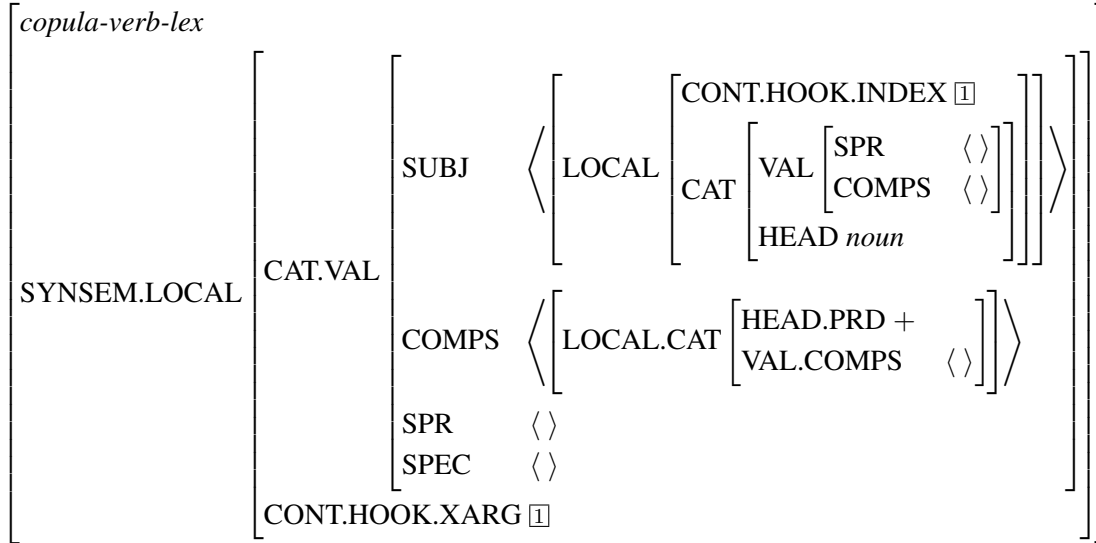
(107) Raising constraints:

$$\left[ \begin{array}{c} \text{SYNSEM.LOCAL.CAT.VAL} \end{array} \left[ \begin{array}{c} \text{SUBJ} \left\langle \left[ \text{LOCAL.CONT.HOOK.INDEX} \quad \boxed{1} \right] \right\rangle \\ \text{COMPS} \left\langle \left[ \text{LOCAL.CONT.HOOK.XARG} \quad \boxed{1} \right] \right\rangle \end{array} \right] \right]$$

Additionally, in the ERG and other DELPH-IN grammars, the feature PRD is used to specify which types can appear as a copula complement, discussed further in §3.3.1. To select the proper

complements, the copula must also constrain its complement to [PRD +]. This results in a type like this:

(108) Basic Copula Lexical Type



A key feature of copulas is that they are semantically empty. In DELPH-IN style HPSG grammars with MRS, this is encoded as not defining a predicate in the relations list of the lexical entry (Copestake and Flickinger 2000). While most lexical entries are specified with a semantic predicate, copulas are not.

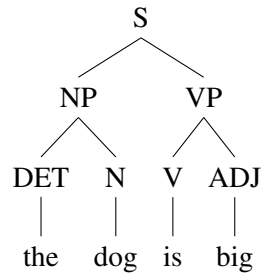
#### 4.2.2 Copula Inflection and Agreement

Copulas tend to inflect for event type features like tense, aspect, or mood (Pustet 2003). This inflection is defined in language-specific definition files with constraints like [TENSE *present*] on the copula's semantic index. Because the copula identifies the semantic index of its complement with its own semantic index, this constraint is also applied to the complement's semantic index. With adjectives, this results in constraints placed on the copula's semantic index also being constrained on the adjective relation in the MRS.

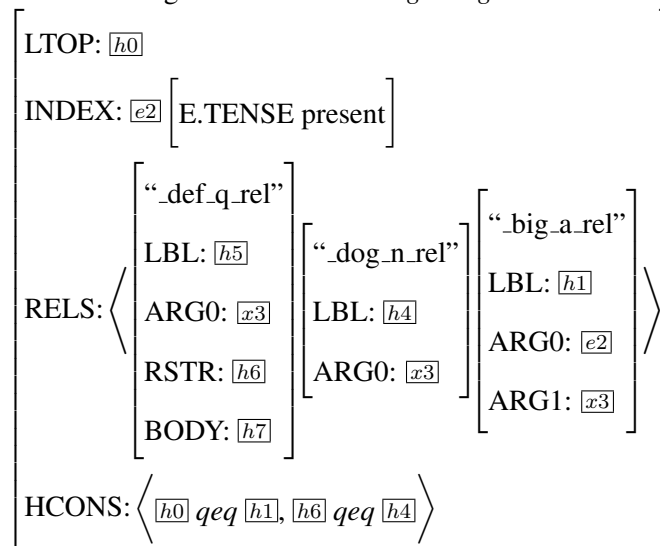
These constraints result in a syntactic tree and associated MRS similar to the following MRS for *the dog is big*. Importantly, note that there is no relation for a copula and the tense value specified on the copula is constrained on the relation for *big*.



- (109) Syntactic tree with copula of English sentence *the dog is big*



- (110) MRS<sup>3</sup> for English sentence *the dog is big*



#### 4.2.3 Copula Complement Selection

Some languages have several copulas which select a different set of complements. In Spanish, the *ser/estar* variation (see e.g. Andrade 1919; Bolinger 1944; Crespo 1946; Bolinger 1947, etc.) could be analyzed as two copulas taking two different (sometimes homophonous) sets of complements. To analyze this, I propose the use of a boolean feature to specify which copula a particular adjective is compatible with. In the case of a more complicated selection process, a more complicated hierarchy could be used to capture this.

The Grammar Matrix customization system is already capable of handling such cases through the use of defining ‘other features’, where the user can define their own syntactic or semantic features

<sup>3</sup>This MRS is created by a grammar created by the final implementation of this analysis, as discussed in Chapter 5

to be used in modeling the lexicon or morphology. Users are able to specify a feature for the copula to constrain on its complement, an example following:

$$(111) \left[ \text{SYNSEM.LOCAL.CAT.VAL.COMPS.FIRST.LOCAL.CAT.HEAD.ADJ\_FORM } \textit{adj1} \right]$$

This can be utilized, along with specifying the *adj-form* on the adjective, to capture copula complement variation.

#### 4.2.4 Summary

I analyze the copula as a semantically empty raising verb, using constraints along the line of the ERG (Flickinger 2000). Copula inflection and agreement features are constrained appropriately, with event features constrained on the copula being properly constrained on its complement by virtue of the copula identifying its complement's semantic index with its own.

### 4.3 Mixed Type Languages

As discussed in §2.4.2, some languages have some sort of split in adjective behavior, which Stassen 2003 names mixed type languages. Some languages are said to be *split*, while others are said to be *switching*.

*Split* type languages have two or more distinct types of adjectives that either have different syntactic behaviors or inflection. These types of phenomena are well suited for analysis by the typed feature structures of HPSG in that different types of adjectives can be defined, with different syntactic or morphological properties, as siblings under a common supertype. This way, shared properties can be pushed up the hierarchy while unique properties can be specified precisely.

*Switching* adjectives are a type of adjective which displays multiple syntactic behaviors or morphological patterns. Switching languages can be categorized into optionally inflecting and optionally copula complement types. This section briefly summarizes the analyses for these behaviors.

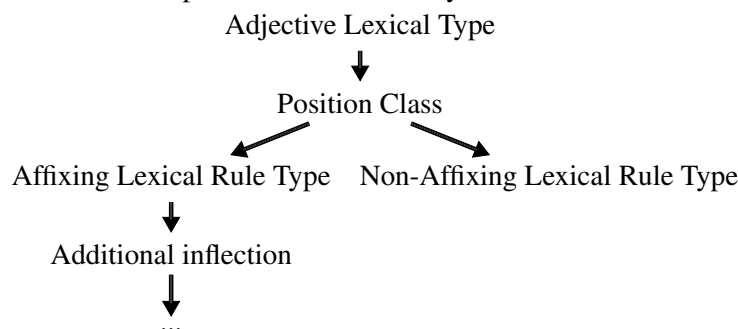
#### 4.3.1 Optional Inflection

Optional Inflection, as discussed further in §2.4.2, is the circumstance in which one adjective appears with two or more different sets of morphological marking. For instance, a predicative adjective

might either be marked for definiteness *or* aspect, usually with slightly different pragmatic or semantic meaning.

I analyze these sorts of bifurcated inflectional patterns with existing Grammar Matrix systems, such as those described in O'Hara 2008 and Goodman 2013. Generally, a position class with at least two lexical rule types is required, where the position class takes the adjective lexical type as its input. Additional position classes necessary for additional inflection then take one of these lexical rule types as their input. If in one case there is inflection, and another case there is not inflection, the same analysis applies, but one lexical rule type has only one lexical rule instance without an affix (l-rule). (112) is a schematic representation of this analysis:

(112) Schematic of Optional Inflection Analysis



#### 4.3.2 *Optionally Copula Complement*

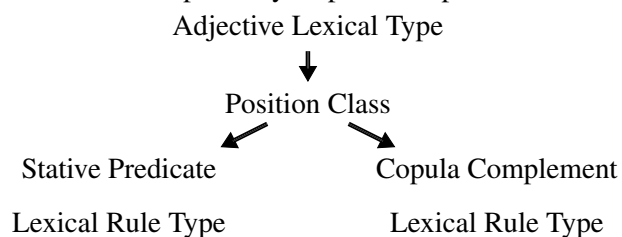
As discussed in §2.4.2, some languages' adjectives can either appear as a copula complement or a stative predicate. I refer to these scenarios as optionally copula complement adjectives. To model these languages, I again utilize the morphological machinery of position classes and lexical rule types, as discussed in §3.2.1, to allow for one lexical type with the two syntactic behaviors, as discussed in §4.1.1.

Informally, this analysis makes use of a lexical type in the lexicon and two separate morphological rules (two separate lexical rule types). The type from the lexicon is underspecified for predicative behavior (copula complement or stative predicate). This lexical type is then the input to both of the competing morphological processes. One of these morphological processes specifies its output to

be a copula complement while the other specifies it to be a stative predicate. Importantly, the output of either of these lexical rule types is not compatible as input to the other.

More formally, a position class with two (or more) lexical rule types takes the adjective lexical type as its input. One of the lexical rule types is constrained to be a stative predicate (specified with the *stative-pred-lex-rule* supertype, shown as a lexical type in (88), and constrained [PRD –]) while the other is constrained to be a copula complement ([PRD +, SUBJ ⟨ ⟩]).

(113) Schematic of Optionally Copula Complement



This way, strings with a copula are licensed by the copula complement lexical rule, while strings without are licensed by the stative predicate lexical rule. Users can specify affixes either directly to these lexical rules or create additional position classes that take either of these lexical rules as input to model additional inflection. By specifying as input one of the two lexical rule types for additional inflection, any additional inflection is therefore only licensed in the proper syntactic construction.

#### 4.4 Constrained Argument Agreement

As detailed in §2.3.2, some languages' adjectives agree in one syntactic construction and not the other. For instance, German adjectives agree with their modificand but not their subject. I analyze this phenomenon as a purely morphological one. In the case of modificand-only agreement, adjectives with inflection are constrained to be attributive only, while adjectives without inflection are constrained to be predicative only. An alternative to this approach, where this bifurcation is in the lexicon, is problematic. In this section, first I further discuss whether this should be considered a lexical or morphological phenomenon in §4.4.1, then I will cover the specific analysis in §4.4.2.

#### 4.4.1 *Lexical or Morphological Process?*

In this subsection, I discuss whether argument constrained agreement is best analyzed as a lexical or morphological phenomenon. I argue that a morphological analysis is best due first to the overt morphological differences and second to specific type hierarchy design considerations. See also §4.1.1.

First, the issue of constrained argument agreement could be considered from a lexical approach. Considering the organization of the lexicon with regard to how these two behaviors are related is an interesting question. An easy answer to this would be to have homophonous lexemes: a lexicon could contain both an attributive-only and a predicative-only version of each adjective. However, this would assert that each adjective in languages that behave this way has two lexical entries, while languages without this behavior have only one.

Alternatively, the morphological approach is natural to this particular phenomenon in that (at least) one instance includes overt morphological marking.<sup>4</sup> Additionally, only one lexical entry per adjective is required because the syntactic bifurcation is left to the morphology.

Considering these positions, it seems natural to consider this process a morphological one.

#### 4.4.2 *Morphological Analysis of Constrained Argument Agreement*

The analysis so far has laid out a set of feature structures in which the external argument of an adjective lexical type is identified with its modificand, and, in stative predicates, the subject. Therefore, a type inheriting from this basic adjective definition with a constrained modificand cannot then forbid the modificand. However, morphological processes are free to output different values than their input. Similarly to the analysis of optionally copula complement adjectives in §4.3.2, my analysis relies on a bifurcated morphological path.

Generally, an underspecified lexical type for the constrained argument agreement adjective is defined without constraints on its syntactic behavior (predicative vs. attributive). This lexical type is then the input for a morphological process with two possible outputs: for modificand-constrained agreement, the process outputs either an inflected lexeme constrained to be attributive only or an

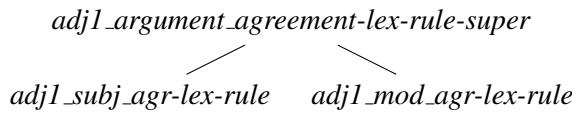
---

<sup>4</sup>This issue would also be of concern with adjectives that agree with their subjects and modificands differently.

uninflected lexeme constrained to be predicative. For predicative-constrained agreement, the predicative only lexeme is inflected and the attributive only lexeme is uninflected.

More specifically, the input lexical type is underspecified for syntactic behavior (predicative vs. attributive). An obligatory position class with two uninflected lexical rule types is created, one lexical rule type specified to be predicative only, the other attributive only. Inflection can be defined on the proper lexical rule type or through an additional position class which takes the proper lexical rule type as input. The predicative only lexical rule also varies between being a copula complement or stative predicate. An example of this position class is below:

- (114) Constrained Argument Agreement Position Class Type Hierarchy:



- (115) Position class defining inputs and marking inflection flag as satisfied

$$\left[ \begin{array}{l} \text{adjl\_argument\_agreement-lex-rule-super} \\ \text{INFLECTED} \left[ \text{ADJ1\_ARGUMENT\_AGREEMENT-FLAG} + \right] \\ \text{DTR adj1-adj-lex} \end{array} \right]$$

- (116) Subject agreement lexical rule constrained to be predicative only (copula complement)

$$\left[ \begin{array}{l} \text{adjl\_subj\_agr-lex-rule} \\ \text{SYNSEM.LOCAL.CAT} \left[ \begin{array}{l} \text{HEAD} \left[ \begin{array}{l} \text{MOD} \langle \rangle \\ \text{PRD} + \end{array} \right] \\ \text{VAL.SUBJ} \langle \rangle \end{array} \right] \end{array} \right]$$

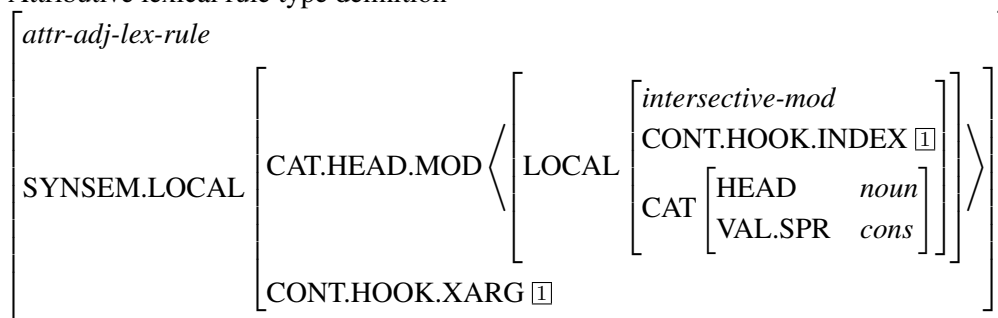
- (117) Modificand agreement lexical rule constrained to be attributive only

$$\left[ \begin{array}{l} \text{adjl\_mod\_agr-lex-rule} \\ \text{SYNSEM.LOCAL} \left[ \text{CAT} \left[ \begin{array}{l} \text{VAL.SUBJ} \langle \rangle \\ \text{HEAD.PRD} - \end{array} \right] \right] \end{array} \right]$$

With this position class, users are able to specify modificand-specific agreement on the *adjl\_mod\_agr-lex-rule* lexical rule or subject specific agreement using the *adjl\_subj\_agr-lex-rule* lexical rule (for more on user interface, see §5.1.3). For more complex agreement patterns, users can define additional position classes that take either the *adjl\_mod\_agr-lex-rule* or *adjl\_subj\_agr-lex-rule* lexical

rule types as input. If the adjective is a stative predicate, the appropriate lexical rule, (116), is defined with the stative predicate lexical rule, shown as a lexical type in (88) (see §4.1.1, §4.3.2 for further discussion). If a lexical rule is attributive, it gets the attributive constraints detailed in §4.1.1, repeated here in its lexical rule form:

(118) Attributive lexical rule type definition

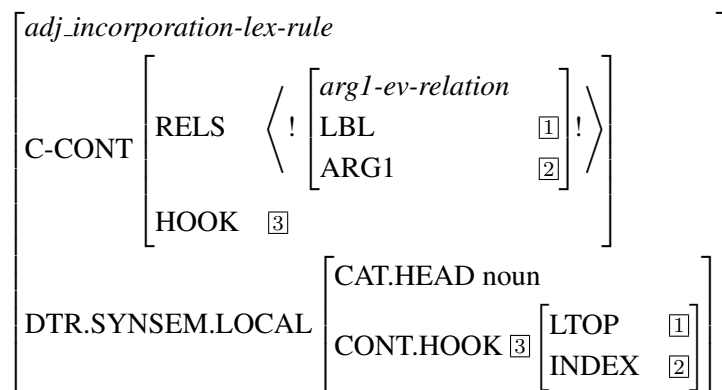


#### 4.5 Morphologically Incorporated Adjectives

As discussed in §2.3.3, attributive adjectives in some languages are affixes of the noun they modify. I analyze these as a morphological process with similar constraints to attributive adjectives. The critical piece of this analysis is that the morphological process also adds a semantic predicate, something typically restricted to lexical entries. Unlike other morphological rules, which just specify an orthographical stem, incorporated adjective morphological rule instances specify both an orthographical stem and a semantic predicate.

Specifically, where adjectives as separate words constrain their modificand, adjectives as incorporated stems constrain the daughter of the lexical rule attaching the adjectival predicate-introducing morpheme. See (119) below.

(119) Incorporated adjective lexical rule type



This lexical rule constrains its daughter to be a noun, similarly to the constraint on adjectives to only modify nouns, and creates a semantic relation with its daughter's LTOP (the pointer to the label of the predication introduced by the noun) as label and identifies its daughter's semantic index as the semantic argument (the ARG1). The semantic predicate and orthographical stem is then specified in a lexical rule instance inheriting from this lexical rule type.

This results in a syntactic tree different than those found in languages with adjectives appearing as words, but an identical semantic representation. A syntactic tree and MRS for Penobscot, a language with incorporated adjectives, is shown in (121) and (122) respectively (data shown in (120)), while the equivalent for English is shown in (123) and (124).

(120) Penobscot incorporated adjective

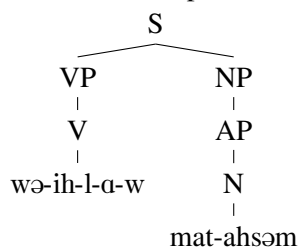
wətihlāw mātahsəmal

wə-ih-l-α-w mat-ahsəm-al

3-tell-NA\_O-DIR-3 bad-dog-OBV

'he tells the bad dog' [aaq-pen] (Quinn 2006)

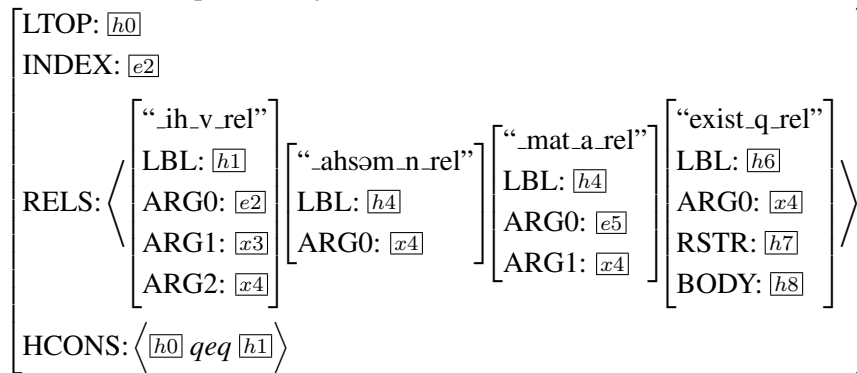
(121) Penobscot incorporated adjective syntactic tree<sup>5</sup>



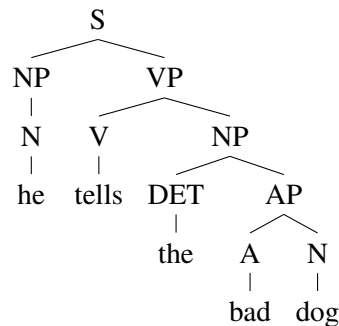
<sup>5</sup>The Penobscot string represented in this syntactic tree is a regularized morphophonological representation, as discussed further in §6.2.1.



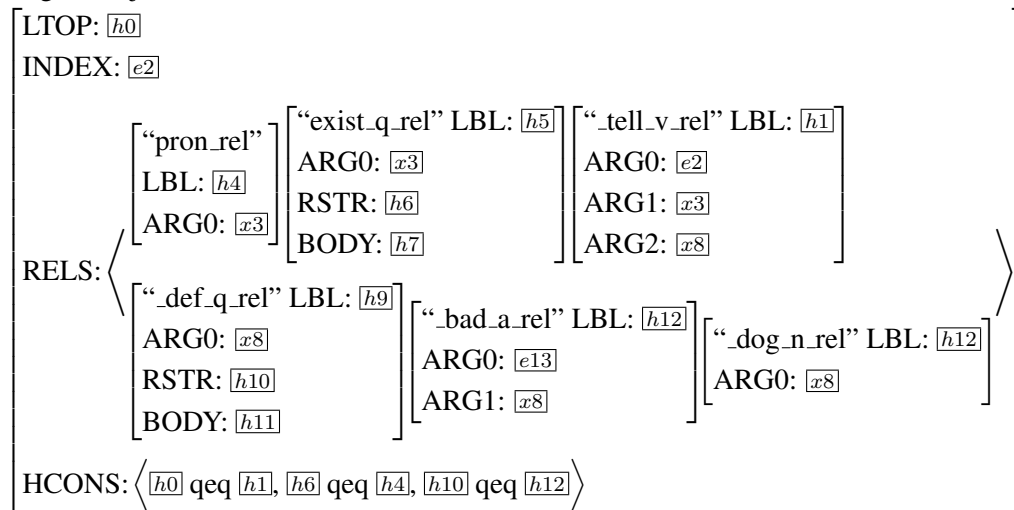
- (122) Penobscot incorporated adjective MRS
- <sup>6</sup>



- (123) English adjective syntactic tree



- (124) English adjective MRS



While the English MRS has several more relations, note the similarity between the verb's relation and the adjective's relation. In the adjectival relations, “\_mat\_a\_rel” in Penobscot and

<sup>6</sup>Because Penobscot has dropped subjects, the Penobscot MRS does not have a semantic relation related to a pronoun or the pronoun's associated quantifier.

“\_bad\_a\_rel” in English, the ARG1 of each points to the appropriate noun: “\_ahsəm\_n\_rel” and “\_dog\_n\_rel”, respectively.

In summary, I propose analyzing incorporated adjectives with similar constraints to attributive adjectives. A morphological rule (lexical rule type) is defined which identifies its input’s (its DTR’s) semantic index to the external argument of the adjective. Affixes (lexical rule instances) can then be defined with an orthographical stem and semantic predicate. This analysis results in similar semantic representations to attributive adjectives with differing syntactic representations.

#### **4.6 Summary**

This chapter has laid out a broad analysis of the behavior of intersective adjectives cross-linguistically. Adjectives are a complex and multifaceted phenomena, and the analysis in this chapter has only begun to analyze the behavior of adjectives. However, some key takeaways from this analysis are:

1. To model adjectives cross-linguistically, it seems necessary to constrain the syntactic behavior of adjectives through morphological processes which apply to lexical definitions.
2. The syntactic behavior of adjectives can be modeled with a three-way split between attributive adjectives, copula complement adjectives, and stative predicate adjectives.

To account for these considerations, I split my analysis into two parts. First is a minimal set of three lexical types stored in the core grammar which define the basic properties of (1) all adjectives, (2) intersective adjectives, and (3) scopal adjectives.<sup>7</sup> Second is a collection of types which are distributed to a user-linguist depending on their choices through the online questionnaire (see §5.2.1 for a discussion of this process). These types can be output either as a lexical type inheriting from one of the core types or as morphological processes taking a core type as input.

While several languages of differing genealogical origin require adjectival syntactic behavior to be defined morphologically as opposed to lexically, I choose to provide user-linguists targeting languages without this need with lexical types providing the appropriate definitions.

---

<sup>7</sup>Note that the constraints included for scopal adjectives are based on previous work included in the Grammar Matrix, and I do not make any assertions to their capacity.

Additionally, I provide an analysis for several additional phenomena, including: copulas, switching adjectives, constrained argument agreement, and morphologically incorporated adjectives. Each of these is designed to complement the core analysis, with switching adjectives and constrained argument agreement wholly relying on this analysis. Additionally, switching adjectives and constrained argument agreement rely on similar bifurcated morphological paths to properly capture the range of phenomena described in Chapter 2.

Chapter 5 provides an overview of the implementation of this analysis, both designing a user interface for creating grammars utilizing these types and an overview of the server-side calculations required to properly assemble these types into a coherent grammar. Chapter 6 provides a numerical evaluation of this analysis over several constructed and natural languages.

## Chapter 5

### IMPLEMENTATION

As discussed in Chapter 4, much of my analysis relies on implementation available as part of the calculations made by the Grammar Matrix customization system. This chapter discusses the specifics of eliciting information regarding the target language by the user-linguist, tools built into the customization system to support the user in language description, and the calculations required to develop output grammars. §5.1 covers the changes to the front end user interface to elicit description and tools to aid description; §5.2 covers the back-end systems to create the output grammar; §5.3 covers the constraints on user description by the system.

To best describe these systems, a review of the Grammar Matrix customization system architecture is required (see §3.1 for additional description). See Figure 5.1 for a schematic representation of the system architecture, which I describe below.

First is what I will refer to as the elicitation loop, where the user interacts with the customization system through a web browser. This interaction is divided into several subpages, many for separate libraries of the customization system, such as Sentential Negation, Argument Optionality, and Information Structure. Much of the linguistic description is done on the Lexicon and Morphology subpages. During the elicitation loop, a single page might have actions executed upon entering some description or making a particular selection, while pages can also interact with each other.

Second, once a user has completed grammatical description and the validation system does not detect any errors, the grammar customization occurs. During this step, the libraries can directly interact with each other. The elicited description is combined and expanded into grammar definitions programmatically without any further user input. The resulting language-specific grammar definitions are delivered to the user along with the static core grammar. When loading this grammar into a parser, these two files interact to parse and generate from data.

Third, during the elicitation loop and immediately prior to the grammar customization, the system executes tests to detect problems. These tests are generally used to forbid input that clashes

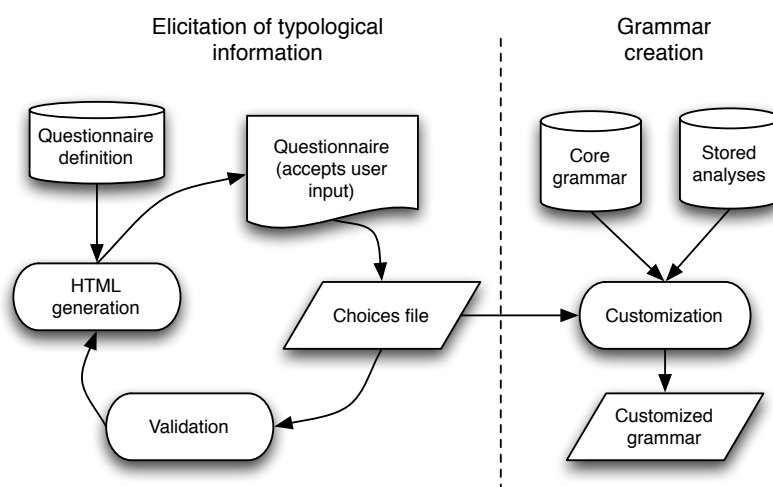


Figure 5.1: Schematic of the Grammar Matrix customization system (Bender et al. 2010)

with the implementation. In addition, I also use these tests in some cases to discourage users from developing analyses I do not think map to any natural language.

This chapter looks at each of these systems in turn.

### 5.1 Extending the Customization System User Interface

In designing grammatical elicitation from the user-linguist, my first consideration was whether or not to devote a new subpage to language-wide description of adjectives, similarly to other libraries of the customization system. However, primarily due to split type languages, I instead leave all grammatical description to the description of adjectival and copula types on the Lexicon page and their lexical rules on the Morphology page. Note that nouns and adpositions also do not have any associated language-wide questions on a separate subpage.

This section is organized as follows. §5.1.1 covers type-specific syntactic behavior of adjectives and copulas. §5.1.2 covers morphological description of adjectives and copulas. §5.1.3 covers the system to enable users to analyze switching type languages.

Additional documentation on using my extension to the customization system is available online.<sup>1</sup>

---

<sup>1</sup>Documentation for the lexicon library is at <http://moin.delph-in.net/MatrixDoc/Lexicon>, while documentation for the morphology library is at <http://moin.delph-in.net/MatrixDoc/Morphology>

### 5.1.1 *Lexicon*

The Lexicon page of the Grammar Matrix customization system is split into different sections for each supported part of speech. Prior to my extensions to the Grammar Matrix, the system supported nouns, verbs, auxiliary verbs, determiners, and case marking and information structure marking adpositions. I have added a section for adjectives and a section for copulas, as seen in Figure 5.2 below.

Importantly, the existing user interface offers users the opportunity to define multiple subtypes of a given lexical type: a user can define lexical types as needed with varying features to capture the facts of their language, some defined with lexical entries and others without. For instance, users are able to specify several types of nouns, some without lexical entries corresponding to gender features, others without lexical entries for number features, and some with lexical entries specified to inherit from the gender and number types. This way, the features are applied properly, but the user is able to capture generalizations about morphology associated with each type on the morphology page.

This sort of type system is critical for capturing split-type languages, where some adjectives behave one way, and others another way. With the type interface built into the Grammar Matrix, users are able to specify multiple types of adjectives while minimizing repetition of constraints.

This section provides an overview of adjective and copula options in turn.

#### *Adjectives*

Like the other parts of speech supported by the customization system, each adjective type defined can have a name and supertype specified. If a user does not specify a supertype, the system calculates the proper supertype given the other choices made below. Users can then specify the syntactic behavior, features, modification direction, and predicative behavior.

Each adjective type can be specified with a syntactic behavior, one of *attributive*, *predicative*, *both*, or *unspecified*. Types specified as either attributive or predicative are defined with the attributive only or predicative only supertype, respectively.

In order to model switching languages and argument agreement languages, that is, phenomena I analyze with morphological rules, the user interface utilizes the *unspecified* option (see §4.1.1 for discussion of my analysis). This option is selected automatically by the system in the case of a user

Figure 5.2: The Lexicon page of the customization system

+>

## Lexicon [\[help\]](#)

On this page you will define **lexical types** and **lexical items** within those types. For most lexical items you must provide both the spelling of the stem and a **predicate** (or **relation**) that identifies the stem's semantic contribution (e.g. `_cat_n_rel` for the noun relation contributed by *cat* or `_sleep_v_rel` for the verb relation contributed by *sleep*). If the predicate is empty upon adding a spelling the page will create a predicate value based on the [DELPH-IN](#) standards. If you completed the Argument Optionality page, be sure to specify whether each verb type allows subject and/or object dropping by selecting the correct value for the `OPT` feature.

▼ all sections

---

### Noun Types | [visualize noun hierarchy](#) (experimental)

Some nouns in this language take adjectives as incorporated affixes: ☐

? ► noun1

Add a Noun Type

---

### Verb Types | [visualize verb hierarchy](#) (experimental)

? ► verb1  
? ► verb2

Add a Verb Type

---

### Adjective Types | [visualize adjective hierarchy](#) (experimental)

Add an Adjective Type

---

## Auxiliary Verb and Copula Types

Auxiliaries may contribute an independent predicate, e.g., English modal *can*. If you define a type that contributes a predicate, you may also specify semantic feature values on the auxiliary, if desired. Alternatively, auxiliaries may contribute no predicate of their own, e.g., English auxiliary *be*. In this case, they do not directly contribute semantic values; instead, they may contribute indirectly by placing constraints on their complements. Note that auxiliary features defined here place no constraints on the semantic values of the complement; constraints on the complement should be defined as complement features.

Add an Auxiliary Type

Define a copula to introduce complements.

[visualize copula hierarchy](#) (experimental)

Add a Copula Type

[Main page](#)  
  
[\\*Gen Info](#)  
[\\*Word Order Number](#)  
[\\*Person Gender](#)  
[\\*Case Dir-inv](#)  
[TAM Features](#)  
[Neg Coord](#)  
[Y/N Qs](#)  
[Info Str](#)  
[Arg Opt](#)  
[?Lexicon Morph](#)  
[Tb Import](#)  
[Test S](#)  
[TbG Options](#)  
  
[Choices file](#)  
 (right-click to download)  
[Save & stay](#)  
[Clear form](#)  
 Create grammar: #  
 tgz, zip

Figure 5.3: Adjectives on the Lexicon page of the customization system

**Adjective Types** | [visualize adjective hierarchy](#) (experimental)

▼ adj1

☐ X **Adjective type 1:**

Type name:

Supertypes:  ▼

Adjectives in constructions are described as **attributive**; those that modify nouns (e.g. "the **big** dog"); and **predicative**, those that predicate their subject (e.g. "the dog is **big**"). Many adjectives in many languages can be both predicative or attributive.

This type of adjective is ☐ attributive ☐ predicative ☐ both ☐ unspecified

Features:

Specify agreement features of the adjective by selecting one of the argument choices; specify inflection for predicative adjectives by selecting a feature to be specified on "the adjective".

This type of adjective behaving attributively modifies items immediately

☐ after the adjective

☐ before the adjective

☐ either position

☐ some other position: this option is not currently supported

☐ This type of adjective behaving attributively must be the only modifier of its modificand

Adjectives of this type behaving predicatively appear as the complement of a copula:

☐ obligatorily

☐ optionally

☐ impossibly



Figure 5.4: Adjective features specification user interface

Features:  
 Specify agreement features of the adjective by selecting one of the argument choices;  
 specify inflection for predicative adjectives by selecting a feature to be specified on "the adjective".

X Name: Gender Value: masculine Specified on: Both positions

Figure 5.5: Modification direction specification user interface

This type of adjective behaving attributively modifies items immediately

☐ after the adjective  
☐ before the adjective  
☐ either position  
☐ some other position: this option is not currently supported

specifying switching or argument agreement.

Prompts are available to the user-linguist to specify features on each sort of lexical type. Features for adjectives come in two categories: agreement features and event features. Features are also said to be *specified on* a particular target. There are four such targets. Agreement features can be specified on *the subject*, *the modified noun*, or *both positions*; event features are specified on *the adjective*. These four options are presented to the user transparently with a note on how to use them. If a user specifies an event feature on an argument position or specifies an agreement feature on the adjective, the user is prompted with an error message instructing them to change their specification.

For instance, if an adjective agrees with the noun it predicates in masculine gender, the user is able to select *gender: masculine* specified on *both positions* as in Figure 5.4.

Languages with adjectives that agree with either only the subject or modified noun, when the user selects either *the subject* or *the modified noun*, the system automatically selects the syntactic behavior *unspecified* and creates a new position class with the two proper lexical rules, one for attributive morphology and another for predicative morphology.

With attributive adjectives, users are able to specify the direction of modification with an option of three different radio buttons as in (5.5): *before the adjective*, *after the adjective*, or *either position*. If an adjective type is defined as unspecified or predicative only and a user selects a modification direction, the user is presented with a warning that the choice will be ignored.

Another option is presented to users, *some other position*, but this position is unselectable and presented with a note informing users the option is not supported. This fourth option is presented to users to make clear this option is not currently implemented. Users could create simplified grammars that do not capture the full array of adjective word order or skip the definition of adjectives all together.<sup>2</sup>

Users defining attributive adjectives are able to define an adjective type with a constraint that the adjective must be the only modifier of its modificand. A checkbox is provided to mark a given type this way (as seen in Figure 5.3).

Lastly, users are able to specify the predicative behavior of adjectives. Similarly to the previous two choices, if a user defines an adjective as attributive or unspecified and then makes a choice for predicative behavior, the system presents a warning that the choice will be ignored. The choices are framed as the availability of an adjective as a copula complement. The adjective can be specified as a complement of a copula *obligatorily*, *optionally*, or *impossibly*. This three way distinction provides users with a way to cleanly specify whether the adjective type is a copula complement, stative predicate, or switching adjective. Optional inflection is handled in the morphology library, discussed in §5.1.2.

### *Copulas*

I have added the option for users to define copulas under the existing heading of *auxiliary verb types*, which has been modified to read *auxiliary verb and copula types*. Like other parts of speech, users can specify a name and any custom supertypes. Users can then specify the *complement type* and any *features* of the copula.

Complement type is displayed as a multi-select dropdown box, where users can select zero to *n* of the options presented in the dropdown. These options are *NPs*, *PPs*, and *APs*. *NPs* and *PPs* are presented as disabled options, designed to indicate to the user that the copula currently only supports *AP* complements, though may support *NPs* and *PPs* in the future.

Like adjectives and the other supported parts of speech, copula types can be defined with fea-

---

<sup>2</sup>Bender 2008 presents an analysis of Wambaya, a radically free order language, utilizing the Grammar Matrix. The foundations of the Grammar Matrix are shown to be useful types for engineering a grammar of a radically free word order language, and I hope my analysis is similarly helpful.

tures. Similarly to adjectives, there are options for *the subject*, *the complement*, and *the copula*. Options for *the subject* are available for referential type features while users can specify event type features on *the copula* and *the complement*. While I have not found any examples of the copula agreeing with its complement, *the complement* is provided as an option for complement selection (see §4.2.3 for more on complement selection).

### *Summary*

This section has described my extensions to the lexicon subpage of the Grammar Matrix customization system. I make two primary additions to this page: the functionality to defining adjective lexical types and copula lexical types. Adjective types are presented with a collection of choices: *syntactic behavior*, *features*, *modification direction*, *unique modification*, and *predicative behavior*. Copula types are presented with a choice of *complement type* and *features*. Beyond those presented here, and unlike other part of speech types, adjectives have several additional user interface elements, which are detailed in §5.1.3.

### *5.1.2 Morphology*

I extend the Grammar Matrix customization system to add support for morphology of adjectives and copulas, as well as adding the capability to define affixes for nouns with adjectival stems. The current section discusses these.

As discussed in §3.2.1, the Grammar Matrix customization system makes a three way distinction within morphology. Users are able to define *position classes*, which can be a prefix or suffix, be marked obligatory, and specify some set of inputs from the defined lexical types, position classes, and lexical rules. Each position class then contains a hierarchy of *lexical rule types*. Each lexical rule type can be specified with a name, supertypes, and features. Finally, each lexical rule type can be specified with *lexical rule instances*, which can be either inflecting or non-inflecting.

I add the capability to the customization system to use these existing user interface options for adjectives and copulas. The only notable modification is that adjective and copula lexical rule types can have features specified on an argument position, congruent with the same ability to specify features on the lexicon subpage, as discussed in the previous subsection. An example of this is seen

in Figure 5.6.

Note that while the customization system allows users to define lexical types for auxiliaries and lexical rules that apply to both main verbs and auxiliaries, I chose to allow for lexical and morphological description of copulas separately from main verbs and auxiliary verbs. This is partially motivated by the linguistic differences between copulas and auxiliary verbs and partially an engineering decision to leave the auxiliary verb library mostly untouched.

### *Incorporated Adjectives*

In order for users to define incorporated adjectival affixes on noun stems, I utilize existing functionality of the customization system to allow users to specify *incorporated stem lexical rule types* to nouns. Because most languages do not require this option, a checkbox is put under the noun section of the Lexicon page which prompts users to activate this feature if their language has adjectives as affixes on nouns. Once this checkbox has been activated, the Morphology subpage displays a button to add an incorporated stem lexical rule type to the Noun section.

Incorporated stem lexical rule types are identical to normal lexical rule types in their presentation on the user interface with the exception that lexical rule instances defined on incorporated stem lexical rule types are able to have a semantic predicate defined alongside an orthographic stem.

### *Summary*

I make three changes to the Morphology subpage of the customization system. First and second are the ability to define morphology on adjective lexical types and copula lexical types, respectively. These changes follow existing functionality of the customization system. Third is the capability to define adjectival incorporated stem lexical rule types on nouns. The user interface to define incorporated adjectives is abstracted from the specifics of the analysis, minimizing the differences between defining incorporated adjective affixes and regular affixes.

#### *5.1.3 Switching*

The analysis of switching constructions relies on lexical types being defined on the lexicon subpage and lexical rules being defined on the Morphology subpage. Users are prompted to complete this in a

Figure 5.6: Morphology User Interface for Adjectives with Agreement Specified

**Adjective Inflection**

▼ adj-number (adj-pc1)

☒ **Adjective Position Class 1:**

Position Class Name:

Obligatorily occurs: ☒

Appears as a prefix or suffix:

Possible inputs:

Morphotactic Constraints:

Lexical Rule Types that appear in this Position Class:

▼ adj-plural (adj-pc1\_lrt1)

☒ **Lexical Rule Type 1:**

Name:

Supertypes:

Features:

Specify agreement features of the adjective by selecting one of the argument choices;  
specify inflection for predicative adjectives by selecting a feature to be specified on "the adjective".

☒ Name:  Value:  Specified on:

Morphotactic Constraints:

Lexical Rule Instances:

☒ Instance 1 ☐ No affix ☒ Affix spelled

Figure 5.7: Option on Lexicon page to activate incorporated stems

**Noun Types** | [visualize noun hierarchy](#) (experimental)

Some nouns in this language take adjectives as incorporated affixes: ☒

Figure 5.8: Noun section of Morphology page with button to add an incorporated stem lexical rule type

▼ noun-pc1

X **Noun Position Class 1:**

Position Class Name:

Obligatorily occurs: ☐

Appears as a prefix or suffix: Prefix

Possible inputs: \*  ▼

Morphotactic Constraints:

Lexical Rule Types that appear in this Position Class:

?

two step process. First, when users either specify that a lexical type is optionally copula complement or specify construction-constrained agreement, they are presented with a note informing them to add additional information to newly created lexical rules on the Morphology subpage.

Second, upon visiting the Morphology page, the user is presented with a pre-populated position class with special lexical rule types offering additional prompts for information. Specifically, these lexical rules can be specified as *attributive*, *predicative*, or *both*; can have the direction of modification specified as *before the adjective*, *after the adjective*, or *either position*; and can be specified as appearing as a copula complement or not (see §5.1.3 for further discussion of these). These options are shown in Figure 5.11.

These three options (syntactic behavior, modification direction, and obligatoriness of a copula) match up with the behavior seen in languages which require these analyses. The system does not present users with the option to make these lexical rules the only modifier of their modificand or define a lexical rule as optionally copula complement because I did not find any instances of this occurring (see §2.3.4 and §2.4.2 for more information of the typology and behavior of unique modification and optionally copula complement adjectives, respectively).

Figure 5.9: Options for copula complement vs. stative predicate

Adjectives of this type behaving predicatively appear as the complement of a copula:

- ☐ obligatorily
- ☐ optionally
- ☐ impossibly

### *Optionally Copula Complement*

On the Lexicon page, users are presented an option to select the obligatoriness of the copula with a given adjective type. To capture optionally copula complement adjectives, an option of *optionally* appearing as a copula complement is presented:

If the user selects *optionally*, the adjective lexical type's syntactic behavior is marked *unspecified* and upon saving the form, a new position class with two lexical rule types is created. The position class's name is created by combining the name of the lexical type with *\_opt\_cop*, and the two lexical rule types are named *\_cop\_comp* and *\_stative\_pred*. The *\_cop\_comp* lexical rule is marked as appearing as a copula complement automatically.

As detailed in §4.3.2, my analysis relies on this bifurcation of the morphological rules to properly license both a copula complement and stative predicate version of an adjective. If there is no affixation associated with these, the user can simply designate both lexical rule types as non-inflecting.

The automatically created lexical rule types are set up properly automatically, such that there are two lexical rule types, one stative predicate and one copula complement. However, any additional lexical rule added to this position class can be defined however the user needs, including features, affixes, etc. For instance, if a language inflects one way in a copula complement construction, another way in a stative predicate construction, and a third way in an attributive construction, users would be able to use this interface to model this hypothetical behavior.

### *Construction-Constrained Agreement*

To model construction-constrained agreement, I continue to use the existing model of features being *specified on* some position, usually arguments or the type being defined. To model construction-constrained agreement, I have designed the questionnaire to offer users to specify a feature on one

Figure 5.10: Automatically created position class with *\_cop\_comp* and *\_stative\_pred* lexical rule types (lexical rule types collapsed)

**Adjective Inflection**

▼ adj1\_opt\_cop (adj-pc1)

**Adjective Position Class 1:**

Position Class Name:

Obligatorily occurs: ☒

Appears as a prefix or suffix: \*

Possible inputs:

Morphotactic Constraints:

Lexical Rule Types that appear in this Position Class:

▶ adj1\_cop\_comp (adj-pc1\_lrt1)

▶ adj1\_stative\_pred (adj-pc1\_lrt2)

of the adjective, the subject, the modified noun, or both positions.

While defining an adjective lexical type on the Lexicon page, when a user selects either of the argument specific choices (*the subject* or *the modified noun*), the system automatically changes the lexical type's syntactic behavior to *unspecified*, displays a message to the user, and upon saving, creates a new position class with the two required lexical rules. These lexical rules function identically to those created with an optionally copula complement adjective type described above. The position class automatically created is named as the combination of the input lexical rule type's name and *\_argument\_agreement*, while the two lexical rule types created are appended with *\_subj\_agr* and *\_mod\_agr*. The lexical rule type named *\_subj\_agr* is specified as predicative only, while the lexical rule type named *\_mod\_agr* is specified attributive only. Choices made regarding modification direction and the obligatoriness of being a copula complement are automatically transferred over to the lexical rule types.



Figure 5.11: Switching lexical rule type options

▼ adj1\_cop\_comp (adj-pc1\_lrt1)

**Lexical Rule Type 1:**

Name:

Supertypes:  ▼

Features:  
Specify agreement features of the adjective by selecting one of the argument choices;  
specify inflection for predicative adjectives by selecting a feature to be specified on "the adjective".

Instances of this Lexical Rule Type can be ☐ attributive ☒ predicative ☐ both

Instances of this Lexical Rule Type behaving attributively modify nouns directly

☐ after the adjective

☐ before the adjective

☐ either position

☐ some other position: this option is not currently supported

? ☒ Instances of this Lexical Rule Type behaving predicatively appear as copula complements

Morphotactic Constraints:

Lexical Rule Instances:

☒ Instance 1 ☐ No affix ☐ Affix spelled

Figure 5.12: Argument specific feature specification with argument choices in dropdown

☒ Name: \*  Value: \*  Specified on ☒

The adjective  
 The subject  
 The modified noun  
 Both positions

This type of adjective behaving attributively modifies items immediately

### *Summary of Switching Implementation*

I utilize automatically generated lexical rule types to analyze optionally copula complement adjectives and construction-constrained agreement adjectives, which I term *switching constructions*. This is done in a two step process: first eliciting choices on the Lexicon page and second describing the specific behavior on the Morphology page. This section has detailed how I have designed the system to elicit this linguistic description from the user-linguist, utilizing a collection of prompts, messages, and pre-population of subpages. When specific choices are made on the Lexicon page, the system automatically generates the proper position class and lexical rule types on the Morphology page. Lexical rule types automatically created by the system in this way have additional prompts, allowing users to select syntactic behavior, modification direction, and a boolean choice about predicative behavior.

#### *5.1.4 Summary*

This section has detailed the extensions to the Grammar Matrix customization system's user interface. My changes extend existing mechanisms on the Lexicon and Morphology subpages to now enable description of adjective and copula lexical types and morphology. Additionally, I developed a user interface for users to describe switching phenomena primarily on the Lexicon page, triggering the pre-population of the required lexical rules on the Morphology page. The following section covers the server-side components of the customization system that take the definitions described in this section as input, and produces a customized grammar of the target language on top of the core definitions.

## ***5.2 The Lexical and Morphological Components of the Customization System***

Once a user has defined adjectives or copulas on the customization system questionnaire (as discussed above in §5.1), the **grammar customization** step is executed to produce an output grammar. This section details the changes I made to this procedure in order to implement the analyses of adjectives and copulas described in Chapter 4, compiling into an output grammar. Because my analysis presented in §4.1.1 relies heavily on the customization system differentiating between a language's required adjectival lexical types vs. adjectival lexical rules, there is a significant amount of calcu-

lation at the grammar customization step, specifically deciding between outputting lexical types or lexical rules.

User specifications are analyzed to create grammar specifications in the DELPH-IN Joint Reference Formalism, a variant of the HPSG formalism. Type description is codified as Type Description Language, or TDL (Copestake 2002). The output grammar is a collection of TDL statements, resulting in a grammar loadable by one of the DELPH-IN processors.<sup>3</sup> An example of TDL is below:

(125) Basic adjective lexical type TDL definition.

```
basic-adjective-lex := norm-sem-lex-item &
  [ SYNSEM [ LOCAL [ CAT [ HEAD adj,
                        VAL [ COMPS < >,
                            SPEC < > ] ] ],
    LKEYS.KEYREL event-relation &
    [ ARG0 #index ] ] ] .
```

Once users have defined lexical types for their target language, the back-end system receives these definitions. These definitions are then used to calculate proper values and assemble TDL definitions of the described types for the output grammar. Each library of the Grammar Matrix is called in sequence: the syntactic libraries are called first, including Sentential Negation, Argument Optionality, and Case; once these other libraries are complete, the system calls the lexicon and morphology library customization scripts. I exclusively extend the lexicon and morphology systems, and do not make any changes to other libraries.

The output is organized into several text files: my changes affect *mylanguage.tdl*, *lexicon.tdl*, *irules.tdl* and *roots.tdl*, where *mylanguage.tdl* is given the name specified by the user-linguist of the target language (e.g. *english.tdl*). *mylanguage.tdl* is where the core language-specific constraints are output, including lexical types, lexical rules, and phrase structure rules, both stored types and types calculated based on user input (e.g. a lexical hierarchy). *lexicon.tdl* stores lexical type instances, including orthographic stems and semantic predicates associated with each instance. *irules.tdl* stores inflecting lexical rule instances (i-rules, see §4.1.2 for more information on i-rules), including orthographic stems and ordering constraints (prefix or suffix); incorporated stem lexical rule instances also have a predicate specified. Non-inflecting lexical rule instances are stored in *lrules.tdl*. *roots.tdl*

---

<sup>3</sup>DELPH-IN processors include the LKB (<http://moin.delph-in.net/LkbTop>), PET (<http://moin.delph-in.net/PetTop>), ACE (<http://moin.delph-in.net/AceTop>), and AGREE (<http://moin.delph-in.net/AgreeTop>)

stores definitions of the initial symbol, or root conditions, for unification. These files are included along with *matrix.tdl*, where the core definitions are stored, and are all loaded together by the processing system used to parse and generate.

This section is structured as followed: §5.2.1 details how lexical type definitions are encoded into TDL. §5.2.2 details the changes and additions to the TDL assembly process for the morphology library with regards to adjectives and copulas.

### 5.2.1 *Lexicon*

This section details the TDL assembly of lexical types for adjectives and copulas.

#### *Adjectives*

I have extended the existing server-side lexicon library to interpret user specifications of adjectives (detailed in §5.1.1) to output TDL specifications of my analysis (described in §4.1). Additionally, the system must detect when a user specifies a particular configuration (such as argument constrained agreement) and output the appropriate TDL. This section details these two topics.

#### *Supertype Validation & Choice Evaluation*

For each adjective, I have designed the the system to evaluate the input choices and calculate supertypes.

If a given type has user-defined supertypes, those choices are formatted into the proper type definition and saved to the output file. If a given type does not have user-defined supertypes, the system assigns it the appropriate supertype for its syntactic behavior: *attr-only-adj-lex* for attributive only adjectives, *pred-only-adj-lex* for predicative only adjectives, and *attr-adj-lex* for *both* adjectives. Additional functionality that is presented in the form of a supertype (e.g. stative predicates with *stative-pred-adj-lex*) is also defined as a supertype of these types, as appropriate. However, in the case that a user-defined type differs from its user-defined supertype's syntactic behavior, the subtype is given the proper supertype (for instance, if the supertype is defined as *both* and the subtype is defined as *attributive only*).

Once the proper supertypes have been calculated, the system evaluates each of the user defined choices: *modification direction*, *unique modification*, and *predicative behavior*. Choices applicable

to only one construction are evaluated only on adjectives with appropriate syntactic behavior defined (e.g. modification direction is only processed on *both* or *attributive only* adjective types).

The choices available to the user for modification direction are *before the adjective*, *after the adjective*, and *either position*. The system does the following actions on the input:

- (126)
1. *before the adjective*: the type is defined with the constraint [POSTHD –] and the syntactic rule type mod-head is added to *mylanguage.tdl*.
  2. *after the adjective*: the type is defined with the constraint [POSTHD +] and the syntactic rule head-mod is added to *mylanguage.tdl*.
  3. *either position*: both mod-head and head-mod is added to *mylanguage.tdl*.

Unique modification constraints are relatively simple. If the user has checked the checkbox specifying that a type must be the unique modifier of its modificand, the proper constraint is output. As discussed in §4.1.2, I utilize the following definition to constrain adjectives to be the only modifier of their modificand, shown in (127).

(127) Unique modification constraint TDL:

```
[ SYNSEM.LOCAL.CAT.HEAD.MOD < [ MODIFIED notmod ] > ].
```

Lastly, choices with regard to predicative behavior are evaluated. Recall that there are three user interface options for predicative behavior: the adjective appears as a copula complement *obligatorily*, *optionally*, or *impossibly*. The behavior for each of these choices is detailed below:

- (128)
1. *obligatorily*: the output type is constrained with [PRD +] and [SUBJ ⟨ ⟩]. If the input is specified to be predicative only, the system adds the predicative only lexical type.
  2. *optionally*: the stative predicate lexical rule is slated to be added to *mylanguage.tdl* and if the input is specified to be predicative only, the system adds the predicative only lexical rule type.
  3. *impossibly*: the output type is constrained with [PRD –], the supertype *stative-pred-adj-lex* is slated to be added to *mylanguage.tdl*, and the output type is given the supertype *stative-pred-adj-lex*. If the input is specified to be predicative only, the system adds the predicative only lexical type.

### *Rule Outputs*

As described above, as the system processes each input adjectival lexical type, it collects a list of lexical types, lexical rule types, and phrase structure rule types to add to *mylanguage.tdl*. The lexical types and lexical rule types, as represented in (102), generally have the same constraints, but are applicable to different types of languages. Therefore, the system outputs the appropriate constraints given user input: for instance, *attr-only-adj-lex* with supertype *attr-adj-lex* for attributive only lexical rule types, applicable to languages with switching constructions. Finally, the system saves the user defined features to *mylanguage.tdl* and each defined stem to *lexicon.tdl*.

### *Copulas*

Because the main focus of this project is adjectives, not copulas, the implementation of copulas is much simpler than the implementation of adjectives. However, the algorithm is largely the same three step process: set up, choice evaluation, and output. First, because my analysis of copula asserts that the copula type is not a useful type in every language for grammatical description, if at least one copula type has been defined, the system outputs the basic copula type definition.

Because the copulas I describe are designed to work with adjectives, the complement of the copula is constrained to be an adjective. However, to support user-linguists targeting languages where the copula takes NP, PP, or sentential complements, I have designed the system to output this constraint on a separate type. This enables users to either easily add additional complement types by hand-editing the TDL or allows users to create additional sibling types if the copula differs somehow for each complement type (for instance, if the PP-introducing complement is different than the AP-introducing complement).

Along with these core definitions, the system adds the boolean feature PRD to the HEAD type in *mylanguage.tdl* using the type addendum syntax:

(129) Head type addendum for PRD feature structure

```
head :+ [ PRD bool ] .
```

Next, the system evaluates the user-specified supertypes of each copula lexical type. If a user does not specify a supertype for a given copula lexical type, the system assigns the the root supertype *adj-comp-copula-verb-lex*, the copula type which constrains its complement's HEAD to be of type

*adj.*

Lastly, the system evaluates the specified features and writes lexical entries. Features specified on *the subject* are constrained at CAT.VAL.SUBJ.FIRST.LOCAL.INDEX while features specified on *the complement* are constrained at CAT.VAL.COMPS.FIRST.LOCAL.INDEX. Features are specified on the lexical types written into *mylanguage.tdl*. Finally, lexical entries are written to *lexicon.tdl*.

### *Summary of Lexicon Library Extension Implementation*

This section has detailed the server-side implementation of my extension to the lexicon library. Both adjectives and copulas are implemented. Both part of speech types are implemented in a similar manner, a three step algorithm that analyzes the choices for types to be included, calculates the required supertypes and feature structures for the output types, and finally outputs the specified features, each lexical type, and each lexical instance. The next section will detail the changes to the morphology library.

#### 5.2.2 *Morphology*

Much of my changes to the morphology library are built on existing systems (O’Hara 2008; Bender et al. 2010; Goodman 2013). Little change has been made to the core functionality of the system, instead utilizing existing server-side systems to process adjective and copula morphology. Minimal changes were made to the back-end system for basic adjective and copula morphology, only modifying the system to process the additional parts of speech.

As described in §5.1.2, the user interacts with three nested sections: *position classes*, *lexical rule types*, and *lexical rule instances*. These are evaluated and translated into the output grammar. *Position classes* and *lexical rule types* are output as lexical rule types in *mylanguage.tdl*, while *lexical rule instances* with affixes defined are output as affixes in *irules.tdl*, those without in *lrules.tdl*. Examples of these from French [fra] are seen in (130)–(133).

(130) defines a position class. Its supertypes constrain the position class’s semantic contribution to be null (*add-only-no-ccont-rule*, [C-CONT ⟨ ⟩]) and enable it to specify inflection (*infl-lex-rule*). The position class is obligatory, and therefore “flips” its inflection flag, constraining it to be [GENDER\_PC-FLAG +].

## (130) Position class type TDL

```
gender_pc-lex-rule-super := add-only-no-ccont-rule & infl-lex-rule &
  [ INFLECTED [ GENDER_PC-FLAG + ],
    DTR adj-lex ].
```

(131) defines a lexical rule type. Its supertype is specified as the position class in (130). This lexical rule type constrains the external argument (XARG) of its input (an adjective) to be *gender: feminine*.

## (131) Lexical rule type TDL

```
feminine_lrt-lex-rule := gender_pc-lex-rule-super &
  [ SYNSEM.LOCAL.CONT.HOOK.XARG.PNG.GEND feminine ].
```

(132) and (133) show an example of a TDL definition of an inflecting lexical rule instance and a non-inflecting lexical rule instance, respectively. The lexical rule instance inherits from a lexical rule type, such as in (131), and optionally specifies an affix (such as in (132)).

## (132) Inflecting Lexical Rule Instance TDL

```
feminine_lrt-suffix :=
%suffix (* e)
feminine_lrt-lex-rule.
```

## (133) Non-inflecting Lexical Rule Instance TDL

```
masculine_lrt-lex := masculine_lrt-lex-rule.
```

To enable morphology to apply to adjectives, choices for adjective morphology now generate position class TDL with the daughter (DTR) specified to be *adj-lex*, as in (130) above. Copula position classes take *cop-lex* as daughter.

*Switching position classes*

There are additional options for each lexical rule type, as detailed in §5.1.3, for the automatically created position classes for switching type adjective morphology. For each lexical rule type, users are prompted to specify the syntactic behavior, modification direction, and ability to be a copula



complement. As described in §5.2.1, these choices are evaluated by the system, adding the appropriate lexical rule types to *mylanguage.tdl*. The implementation of switching adjectives relies on these two processes: first, the pre-population of the morphology subpage with additional user interface options for syntactic behavior, modification direction, etc.; second, the addition of the appropriate lexical rules to *mylanguage.tdl* depending on choices made on the lexicon subpage.

As described in §4.1.1, the choices on switching lexical rule types produce similar constraints as choices on lexical types. Attributive only adjective lexical rule types are constrained to be [SUBJ  $\langle \rangle$ , PRD –], predicative only adjective lexical rule types are constrained to be [MOD  $\langle \rangle$ ], adjective lexical rule types that can be both are unconstrained. Lexical rule types specified as copula complements are constrained [PRD +], others constrained [PRD –]. Modification direction is constrained on lexical rule types with identical constraints to lexical types: adjectives appearing after the noun they modify are constrained [POSTHD +], adjectives appearing before the noun they modify are constrained [POSTHD –], adjectives that are licensed on either side are underspecified.

The following is a sample of the lexical rule types produced through this system:

(134) Switching position class TDL

```
switch_opt_cop-lex-rule-super := add-only-no-ccont-rule &
  const-lex-rule &
  [ INFLECTED.SWITCH_OPT_COP-FLAG +,
    DTR switch-adj-lex ].
```

(135) Copula complement lexical rule type TDL

```
switch_cop_comp-lex-rule := switch_opt_cop-lex-rule-super &
  [ SYNSEM.LOCAL.CAT [ HEAD [ MOD < >,
                                PRD + ],
                                VAL.SUBJ < > ] ].
```

(136) Stative predicate lexical rule type TDL

```
switch_stative_pred-lex-rule := switch_opt_cop-lex-rule-super &
  stative-pred-lex-rule &
  [ SYNSEM.LOCAL.CAT.HEAD.MOD < > ].
```

### *Incorporated Stem Lexical Rule Type*

My analysis of incorporated stems requires little change to the customization system. As described in §5.1.2, users are able to specify predicates along with stems for incorporated lexical rule types. Predicates for lexical rule types are stored along with the orthographical stems in new data structure

which represents each lexical rule instance. This way, stems and, when appropriate, predicates, are stored together. The system adds *incorporation-lex-rule* (from (119)) to *mylanguage.tdl* and adds the actual lexical rule instance to *irules.tdl*, both stem and predicate, as in (137):

(137) Incorporation lexical rule instance TDL from Penobscot [aaq-pen]

```
mat-lrtl-prefix :=
%prefix (* mat)
mat-lrtl-lex-rule &
[ C-CONT.RELS.LIST.FIRST.PRED "_mat_a_rel" ].
```

### *Summary of Morphology Library Extension Implementation*

This subsection has covered my extensions to the server-side morphology library of the Grammar Matrix customization system. Adjectives and copulas have been added to the system, utilizing existing analyses of morphology to model the new parts of speech. Switching position classes, created automatically by the lexicon user interface subpage, have additional choices which are evaluated and produce constraints on the specified lexical rule types. Incorporated stem lexical rule types are specified with an orthographical stem and semantic predicate, outputting a lexical rule with a semantic predicate in *irules.tdl*.

#### 5.2.3 *Summary*

The section has covered my changes to the Grammar Matrix grammar customization mechanism. Changes were made to the existing lexicon and morphology libraries, adding the capability to define adjective and copula types as well as morphology for each of these part of speech types. Changes to the lexicon library include a three step process for each type, making basic calculations based on inputs, validation and calculating supertypes, and then outputting required types and lexical rules. Changes to the morphology library include adding the capability to define copula and adjective morphology, switching position classes, and incorporation. The next section will detail the changes to the validation component of the customization system.

### 5.3 *Constraining the Customization System*

The Grammar Matrix customization system provides a vast range of possible grammars. However, many of these grammars are thought to not map to any natural language. The Grammar Matrix therefore includes a **validation system**, designed to provide feedback to users when they select a set of choices that is thought to be not viable or is not yet analyzable. The validation system is also designed to reduce user mistakes and enforce required specifications. This section provides a brief overview of my extension to the validation component.

My extension to the validation component primarily attempts to reduce user errors and remind users to fill out required parts of the questionnaire. However, a few key constraints are also built on linguistic generalizations. This section will briefly detail these two categories.

#### 5.3.1 *Encoding Linguistic Generalizations*

I extend the validation system to forbid one position class from having both incorporated stem lexical rule types and normal lexical rule types. This is due to Algic languages such as Penobscot discussed in §2.3.3, where adjectival affixes only appear in minimal pairs with other adjectival affixes as opposed to other syntactic affixation. If users define a single position class with both incorporated lexical rule types and regular lexical rule types, they are presented with an error. This isn't a perfect solution, as the system is flexible enough to avoid this validation, but it provides a certain level of error reduction.

#### 5.3.2 *Reducing User Errors*

Because my implementation of switching adjectives relies on pre-population of segments of the questionnaire, it is important to guide users to fill in this pre-populated skeleton. It is possible that a user defines a switching adjective on the Lexicon page and then immediately requests grammar customization. However, because validation is performed before grammar customization, this situation results in a user being presented with a validation warning message.

Similarly, it is possible to define a lexical type or switching lexical rule type as obligatorily or optionally appearing as a copula complement without defining a copula. In this case, the user is presented with a warning message that their types will not be usable without a copula being defined.

Lastly, because users can specify supertypes for both lexical types and lexical rule types, I have extended to the validation system to present an error if a specified supertype's choices or features conflict with the type's choices. Importantly, I have enabled a sort of unification for this constraint. For instance, if a user specifies an adjective lexical type as appearing *either before or after its modificand*, a subtype of this lexical type can then constrain its modification direction to *before*, *after*, or *either*. Alternatively, if an adjective lexical type is specified as appearing *before its modificand*, it cannot be a supertype or input to a type specifying a modification direction other than *before*. This allows for maximum exploitation of multiple inheritance while enforcing basic constraints.

#### **5.4 Summary**

This chapter has detailed my implementation across the three key systems of the Grammar Matrix: the user interface, the server-side libraries, and the validation system. In each case, I extended two existing libraries: the lexicon library and morphology library. The user interface was extended to allow user-linguists to define adjective and copula types. The server-side libraries were extended to enable description of adjectives and copulas. The validation system was adapted to require particular choices on both adjectives and copulas, while also disallowing particular combinations of choices I believe to not represent a natural language.

The next chapter will detail an evaluation of my implementation and analysis.

## Chapter 6

### EVALUATION

This chapter presents an evaluation, based on the framework developed in Drellishak 2009; Saleem 2010; Crowgey 2012 and Song 2013. First, I present a set of grammars describing delexicalized and simplified representations of the syntactic phenomena I have developed an analysis for, which I refer to as **pseudo-languages**. Second, I present a set of grammars describing natural languages. Each of these grammars is created by my system and tested against a set of strings. Each test contains a choices file, the saved choices from the customization system, a grammar output from the customization system, and a set of intended grammatical and ungrammatical strings. §6.1 covers the pseudo-languages I created, while §6.2 covers the natural languages used in evaluation.

#### **6.1 Pseudo-Languages & Regression Tests**

The pseudo-languages used for this evaluation are hand-made representations of the phenomena I have analyzed. Each pseudo-language consists of a simple lexicon representing the part of speech the string refers to, such as *n* for noun, *tv* for transitive verb, *adj* for adjective, and so on. Only choices relevant to a specified phenomena are described. This way, each pseudo-language and accompanying test suite is intended to be a minimal test of each feature of the customization system. These test not only the analysis but also the implementation, and subsequently serve as tests for future development to avoid system regressions.

I have added 65 pseudo-language test suites to the regression testing system. Many of these are very similar, and do not need to be individually detailed. Generally, the tests are divided into eight groups of tests. This section details each group in §6.1.1.

##### *6.1.1 Pseudo-Languages*

The pseudo-languages can be categorized into eight distinct groups:

1. Basic attributive

2. Word order
3. Basic both
4. Argument constrained agreement
5. Split type languages
6. Stative predicate
7. Switching type languages
8. Copula

Each of these sections has several pseudo-languages, depending on the combinatorial potential of the phenomena involved. Generally, each pseudo-language only pivots on one combinatorial factor, except when two or more factors are thought to interact in interesting ways.

Each pseudo-language test consists of a choices file, test suite, and gold semantic representations. For instance, one of the smallest pseudo-languages is *adj\_n\_infl*, which tests adjectival inflection and word order. The test suite is as follows:<sup>1</sup>

(138) Test suite for inflecting stative predicate adjective pseudo-language:

```
adjpres n
adjpast n
*n adjpres
*n adjpast
*adj n
*n adj
```

The language describing this set of strings is described by these choices (simplified to the relevant adjective choices on the Lexicon and Morphology subpages):

(139) Relevant choices for inflecting stative predicate adjective pseudo-language:

```
section=lexicon
noun1_det=opt
noun1_stem1_orth=n
noun1_stem1_pred=_n_n_rel
verb1_name=tv
verb1_valence=trans
```

---

<sup>1</sup>The asterisk \* is used to mark a negative test instance.

```

    verb1_stem1_orth=tv
    verb1_stem1_pred=_tv_v_rel
verb2_name=iv
verb2_valence=intrans
    verb2_stem1_orth=iv
    verb2_stem1_pred=_iv_v_rel
adj1_name=adj
adj1_mod=pred
adj1_predcop=imp
    adj1_stem1_orth=adj
    adj1_stem1_pred=_adj_a_rel

section=morphology
adj-pcl_name=adj_tense
adj-pcl_obligatory=on
adj-pcl_order=suffix
adj-pcl_inputs=adj
    adj-pcl_lrt1_name=adj_pres
        adj-pcl_lrt1_feat1_name=tense
        adj-pcl_lrt1_feat1_value=present
        adj-pcl_lrt1_feat1_head=adj
        adj-pcl_lrt1_lril_inflecting=yes
        adj-pcl_lrt1_lril_orth=pres
adj-pcl_lrt2_name=adj_past
    adj-pcl_lrt2_feat1_name=tense
    adj-pcl_lrt2_feat1_value=past
    adj-pcl_lrt2_feat1_head=adj
    adj-pcl_lrt2_lril_inflecting=yes
    adj-pcl_lrt2_lril_orth=past

```

This choices file then generates a grammar in TDL, which is loadable by the DELPH-IN processing tools such as ACE, which outputs the following MRS for the string *adjpast n*:

(140) MRS for pseudo-language string *adjpast n*:

$$\left[ \begin{array}{l} \text{LTOP: } [h0] \\ \text{INDEX: } [e2] \left[ \begin{array}{l} \text{E.TENSE: past} \end{array} \right] \\ \text{RELS: } \left\langle \begin{array}{l} \left[ \begin{array}{l} \text{"_adj_a_rel"} \\ \text{LBL: } [h1] \\ \text{ARG0: } [e2] \\ \text{ARG1: } [x3] \end{array} \right] \left[ \begin{array}{l} \text{"_n_n_rel"} \\ \text{LBL: } [h4] \\ \text{ARG0: } [x3] \end{array} \right] \left[ \begin{array}{l} \text{"exist_q_rel"} \\ \text{LBL: } [h5] \\ \text{ARG0: } [x3] \\ \text{RSTR: } [h6] \\ \text{BODY: } [h7] \end{array} \right] \end{array} \right\rangle \\ \text{HCONS: } \left\langle [h0] \text{ qeq } [h1] [h6] \text{ qeq } [h4] \right\rangle \end{array} \right]$$

These MRS are hand inspected to ensure they represent the analysis I intended. Once an MRS has been vetted, I add it to the regression test suite as a gold-standard to detect new system errors associated with future development.

The following subsections will briefly detail each of these eight groups. Note that each pseudo-language is prefixed with *adj-* to simplify user interaction with the test suite.

### *Basic attributive*

This collection of pseudo-languages tests the basic functionality of attributive adjectives and incorporated adjectives. This group includes twelve pseudo-languages. Pseudo-languages with varying modification direction, number of modifiers, and agreement are created, representing six pseudo-languages. The other six pseudo-languages represent incorporation, which vary in the order of the affix (prefix or suffix), number of possible affixes, and agreement.

The test suite for the pseudo-language *adj-adj-n*, a basic attributive pseudo-language, is shown in Figure 6.1.



|                  |                      |                      |                      |
|------------------|----------------------|----------------------|----------------------|
| adj1 n iv        | adj1 adj1 n iv       | n tv adj1 n          | n tv adj1 adj1 n     |
| adj1 n tv n      | adj1 adj1 n tv n     | adj1 n tv adj1 n     | n adj1 iv            |
| n adj1 adj1 iv   | n tv n adj1          | n tv n adj1 adj1     | n adj1 tv n          |
| n adj1 adj1 tv n | n adj1 tv n adj1     | n adj1 tv adj1 n     | adj1 n tv n adj1     |
| adj1 n adj1 iv   | n cop adj1           | n adj1 cop adj1      | n adj1 adj1 cop adj1 |
| adj1 n cop adj1  | adj1 adj1 n cop adj1 | adj1 n adj1 cop adj1 | *n adj1 cop          |
| n adj1 adj1 cop  | *adj1 n adj1 cop     | *cop adj1 n          | *adj1 cop n          |
| n adj1           | *adj1 n              | *adj1 n adj1         | *adj1 adj1 n adj1    |
| adj1 adj1 n      | *adj1 adj1 adj1 n    | *n adj1 adj1         | *n adj1 adj1 adj1    |
| adj1 n cop       | *cop n adj1          | *n cop adj1 adj1     | *n iv adj1           |
| adj1 iv n        | *adj1 iv             | *adj1 cop            | *cop adj1            |
| iv adj1          | *n tv adj1           |                      |                      |

Figure 6.1: Test suite for pseudo-language *adj-adj-n**Basic both*

This collection of pseudo-languages tests the basic interaction between the modification direction choices and two predicative behavior choices: stative predicates and copula complements. This results in six pseudo-languages: *adj-both-either-cop*, *adj-both-either-stative*, *adj-both-post-cop*, *adj-both-post-stative*, *adj-both-pre-cop*, *adj-both-pre-stative*.

The test suite for the pseudo-language *adj-both-either-cop*, a pseudo-language representing the phenomena of both attributive and predicative adjectives with the attributive adjective appearing either pre-head or post-head and the predicative adjective appearing as a copula complement, is shown in Figure 6.2.

*Stative predicate*

This collection of pseudo-languages tests word order, subject agreement, and event inflection (tense, aspect, mood, etc.), all in the context of stative predicates. There are six pseudo-languages in this group. Tense is used as a prototypical event inflection. Note that the stative predicate order relies on the word order specified on the Word Order subpage, such that the VSO choice results in AS order, the SVO choices results in the SA order, etc.

The test suite for the pseudo-language *adj-adj-n*, a basic stative predicate pseudo-language, is

|                  |                      |                      |                      |
|------------------|----------------------|----------------------|----------------------|
| adj1 n iv        | adj1 adj1 n iv       | n tv adj1 n          | n tv adj1 adj1 n     |
| adj1 n tv n      | adj1 adj1 n tv n     | adj1 n tv adj1 n     | n adj1 iv            |
| n adj1 adj1 iv   | n tv n adj1          | n tv n adj1 adj1     | n adj1 tv n          |
| n adj1 adj1 tv n | n adj1 tv n adj1     | n adj1 tv adj1 n     | adj1 n tv n adj1     |
| adj1 n adj1 iv   | n cop adj1           | n adj1 cop adj1      | n adj1 adj1 cop adj1 |
| adj1 n cop adj1  | adj1 adj1 n cop adj1 | adj1 n adj1 cop adj1 | *n adj1 cop          |
| n adj1 adj1 cop  | *adj1 n adj1 cop     | *cop adj1 n          | *adj1 cop n          |
| n adj1           | *adj1 n              | *adj1 n adj1         | *adj1 adj1 n adj1    |
| adj1 adj1 n      | *adj1 adj1 adj1 n    | *n adj1 adj1         | *n adj1 adj1 adj1    |
| adj1 n cop       | *cop n adj1          | *n cop adj1 adj1     | *n iv adj1           |
| adj1 iv n        | *adj1 iv             | *adj1 cop            | *cop adj1            |
| iv adj1          | *n tv adj1           |                      |                      |

Figure 6.2: Test suite for pseudo-language *adj-both-either-cop*

shown in Figure 6.3.

|           |            |            |
|-----------|------------|------------|
| adj n     | *n cop adj | *adj n cop |
| n adj     | *cop adj n | *adj cop n |
| n adj cop | *cop n adj |            |

Figure 6.3: Test suite for pseudo-language *adj-adj-n*

### Word order

This collection of pseudo-languages tests the interaction between my modifications to the lexicon and morphology libraries and the word order library. While most of the other pseudo-languages use SVO word order, this collection of pseudo-languages tests that the verbal word order systems interact properly with the adjective-modificand word order systems. Specifically, the tests cover the functionality of adjectives in OVS, OSV, SOV, VOS, and VSO languages.

The test suite for the pseudo-language *adj-osv*, a pseudo-language that tests the interaction between OSV verbal word order and attributive adjectives appearing pre-head, is shown in Figure 6.4.

|                |                 |                 |                |
|----------------|-----------------|-----------------|----------------|
| n n tv         | adj n n tv      | n adj n tv      | adj n adj n tv |
| n n tv adj     | *n n adj tv     | *n tv n         | *n tv adj n    |
| adj n tv n     | *adj n tv adj n | *n tv n adj     | *n adj tv n    |
| n adj tv n adj | *adj n tv n adj | *n adj tv adj n | *tv n n        |
| tv adj n n     | *tv n adj n     | *tv adj n adj n | *adj tv n n    |
| tv n n adj     |                 |                 |                |

Figure 6.4: Test suite for pseudo-language *adj-osv*

### *Argument Constrained Agreement*

This collection of pseudo-languages tests argument constrained agreement. These are split into two equal sets, each with four pseudo-languages. These pseudo-languages have two sets, one which tests agreement constrained to the subject, and another set which tests agreement constrained to the modificand. These are similar to the *both* set, varying in modification direction and predicative behavior.

The test suite for the pseudo-language *adj-infl-mod\_post\_stative*, a language which tests agreement constrained to the modificand with post-head attributive adjectives and stative predicate predicative adjectives, is shown in Figure 6.5.

|            |             |             |             |
|------------|-------------|-------------|-------------|
| nm adj     | nf adj      | nm adjm iv  | nf adjf iv  |
| adjm nm iv | *adjf nf iv | *adjf nm iv | *adjm nf iv |
| adj nm iv  | *adj nf iv  | *nm adjm    | *nf adjf    |
| nf adjm    | *nm adjf    | *nf adjm iv | *nm adjf iv |
| nm adj iv  | *nf adj iv  |             |             |

Figure 6.5: Test suite for pseudo-language *adj-infl-mod\_post\_stative*

### *Split-type languages*

This collection of pseudo-languages tests the interaction between multiple adjective types defined, varying on syntactic behavior, predicative behavior, and split inflection versus split copula complements. This results in eight pseudo-languages.

The test suite for the pseudo-language *split\_adj\_cop\_n\_adj*, a pseudo-language that exemplifies

the phenomenon of one set of adjectives appearing as stative predicatives and another set appearing as copula complements, is shown in Figure 6.6.

|                 |                  |                  |                  |
|-----------------|------------------|------------------|------------------|
| n adj1 cop      | n adj2           | *n adj2 cop      | *n adj1          |
| adj1 n          | *adj2 n          | *n cop adj1      | *cop adj1 n      |
| cop n adj1      | *adj1 n cop      | *adj1 cop n      | *n cop adj2      |
| cop adj2 n      | *cop n adj2      | *adj2 n cop      | *adj2 cop n      |
| n adj1-infl     | *adj1-infl n     | *n adj2-infl     | *adj2-infl n     |
| n adj1-infl cop | *cop adj1-infl n | *cop n adj1-infl | *adj1-infl n cop |
| adj1-infl cop n | *n cop adj2-infl | *cop adj2-infl n | *cop n adj2-infl |
| adj2-infl n cop | *adj2-infl cop n |                  |                  |

Figure 6.6: Test suite for pseudo-language *split\_adj\_cop\_n\_adj*

### *Switching-type languages*

This collection of pseudo-languages tests the switching-type adjective system. There are five pseudo-languages, two for optional inflection and two for optionally copula complement adjectives, with one testing optional inflection on optionally copula complement adjectives.

The test suite for the pseudo-language *switching\_infl\_cop\_infl*, a pseudo-language representing an optionally copula complement adjective type where inflection appears on the copula in copula complement constructions and on the adjective in stative predicate constructions, is shown in Figure 6.7.

|               |                |                |                |
|---------------|----------------|----------------|----------------|
| n copinfl adj | n adjinfl      | *n cop adj     | *n adj         |
| adj n         | *adjinfl n     | *n adjinfl cop | *cop adjinfl n |
| cop n adjinfl | *adjinfl n cop | *adjinfl cop n | *n adj copinfl |
| copinfl adj n | *copinfl n adj | *adj n copinfl | *adj copinfl n |
| n adj cop     | *cop adj n     | *cop n adj     | *adj cop n     |
| adj n cop     |                |                |                |

Figure 6.7: Test suite for pseudo-language *switching\_infl\_cop\_infl*

### *Copula*

This set of pseudo-languages contains eleven pseudo-languages and tests the word order between copula and complement, agreement and inflection of the copula, and split copula examples. The split copula examples constitute three pseudo-languages, differing in the word order of copula and complement in a copula complement construction versus subject and adjective in a stative predicate construction. Also tested are two copula types where one type of copula takes one set of adjectives while another copula type takes another set of adjectives. This set of languages also includes tests on word order and copula agreement and tense inflection.

It is important to note that this set of pseudo-languages does not contain an example where the order of adjective and subject in stative predicate constructions differs from the order of subject, copula, and adjective in copula complement languages or the order of subject and verb. As noted in §4.1.2, this combination is not supported by the word order library of the customization system due to the need to define multiple word orders.

The test suite for the pseudo-language *cop\_adj\_agr*, a pseudo-language representing an adjective agreeing with its subject in a copula complement construction, is shown in Figure 6.8.

|             |              |              |              |
|-------------|--------------|--------------|--------------|
| nm cop adjm | nf cop adjf  | *nm cop adjf | *nf cop adjm |
| nm adjm cop | *cop adjm nm | *cop nm adjm | *adjm nm cop |
| adjm cop nm | *nf adjf cop | *cop adjf nf | *cop nf adjf |
| adjf nf cop | *adjf cop nf | *nm adjm     | *adjm nm     |
| nf adjf     | *adjf nf     | *n adj       | *n adjm      |
| nm adj      |              |              |              |

Figure 6.8: Test suite for pseudo-language *cop\_adj\_agr*

### 6.1.2 Summary

This section has summarized the pseudo-language test suites I use to test the functionality of my extensions to the Grammar Matrix. I have added 65 pseudo-language tests, which can be clumped into eight groups.

## 6.2 *Natural Languages*

In addition to pseudo-languages, I test my analysis and implementation on eleven natural languages. These are divided into two equal groups: the first group is illustrative languages, including those discussed in Chapter 2 and those used in aiding development of the system; the second group is held out languages, none of which directly considered during the development of the system.

While the pseudo-language test suites are comprised of hand-crafted strings and grammaticality judgments, these natural language test suites are different. Each natural language test suite is comprised of a set of strings from a source, most of which are grammatical, along with a hand-crafted set of ungrammatical strings. These test suites help determine if my system is adequate in creating grammars which correctly distinguish between grammatical and ungrammatical strings.

Because I do not have gold standard syntactic or semantic representations, I am the sole judge of the veracity of the system output syntactic and semantic representations. I base these judgments on the description of phenomena in the source document and how I analyze these phenomena. Example sentences and associated grammaticality judgments for each are gathered from various sources, as cited. Some of these languages are part of Language CoLLAGE, a collection of Grammar Matrix grammars, test suites, and choices files (Bender 2014).

The goal of developing these test suites is to find the simplest examples exemplifying the adjective phenomena. However, this is sometimes difficult. In curating test suites, I try to avoid examples with phenomena unrelated to adjectives. If a given phenomenon interacts with adjectives somehow, it is modeled, otherwise, it is not. If the Grammar Matrix customization system doesn't currently support some interacting phenomenon, I try to simplify examples as little as possible to maintain the representation of the behavior of the language while allowing for adjectives to be modeled sufficiently. Because the Grammar Matrix doesn't support complex morphophonology, in some cases, I utilize regularized representations of the morphosyntax. These representations are generally based on morpheme segmentation found in an IGT format in source texts.<sup>2</sup>

Furthermore, in curating test suites, while verbatim examples from texts are preferred, and native speaker judgments most preferred, I utilize many constructed examples, that is, examples that I have

---

<sup>2</sup>It is possible to integrate a morphophonological processor with Grammar Matrix grammars, as in Bender and Good 2005; Crowgey 2014.



Figure 6.9: General geographic distribution of illustrative languages

constructed based on the data and the prose description of phenomena found in the cited work and grammars. These constructed examples are primarily of ungrammatical strings, but also to simplify complex examples. These are constructed with utmost care, and are designed to as accurately as possible represent the data described in the descriptive work, but there are sure to be assumptions made on my part that do not represent the language properly. The full test suites are freely available for downloading along with the Grammar Matrix.<sup>3</sup> Any mistake is my own.

### 6.2.1 *Illustrative Languages*

I used six illustrative languages to aid development. While these languages do not represent the most diverse set of languages, each of these languages exhibits a different set of phenomena. These are detailed below:

1. English [eng] (Germanic, Indo-European): multiple adjective types with different syntactic behavior.

---

<sup>3</sup>Test suites available at <http://www.trimbleworks.us/linguistics/> or as part of the SVN repository at <svn://lemur.ling.washington.edu/shared/matrix/trunk>. Instructions for downloading at <http://www.delph-in.net/matrix/>

2. French [fra] (Romance, Indo-European): multiple adjective types with different modification directions, including either direction.
3. Penobscot [aaq-pen] (Algonquian, Algic): morphologically incorporated adjective stems on nouns.
4. Maori [mri] (Oceanic, Austronesian): optionally copula complement switching language with at most one attributive adjective per noun.
5. German [deu] (Germanic, Indo-European): construction-constrained agreement between adjectives and nouns.
6. Russian [rus] (Slavic, Indo-European): stative predicate in present tense, copula complement in other tenses.

The geography of these is shown in Figure 6.9. This section will detail the analysis of each of these.

### *English [eng]*

English adjectives are notable for two reasons. First, they do not take any agreement morphology and always appear before the noun they modify. Second, while most English adjectives can be either attributive or predicative, there is a sizeable group of adjectives that can only be attributive and another group that can only be predicative. English also exhibits a fairly simple copula.

To test English, a test suite with 57 items was created, with 13 grammatical examples and 44 ungrammatical examples. Data was gathered from native speakers, including the author,<sup>4</sup> as well as from examples included with the English Resource Grammar (Flickinger 2000). The test suite includes basic instances of attributive and predicate adjectives, along with multiple adjectives and copula agreement tests. For instance:

- (141) the black dog barks  
       \*the dog big barks  
       the mock interview fails  
       \*the interview is mock  
       the dog is awake  
       \*the awake dog barks  
       I am big  
       \*you am big

---

<sup>4</sup>Thanks to Cailin Swarm for additional native speaker judgments.



The English grammar models tense, person, and number, ignoring other phenomena when not directly applicable to analyzing grammatical sentences.

### *French [fra]*

French adjectives are of interest because there are three distinct types of adjective with varying word order. The first type is regular adjectives, which appear post-head; the second type is the so-called BAGS adjectives (Knop 1971), which appear pre-head; the third type is a set sometimes referred to as figurative adjectives, which can appear either before or after the adjective they modify with minor semantic differences. French adjectives also agree in number and gender with either their subject or modificand. French copulas also agree with the subject in person and number, and inflect for tense and aspect.

The French test suite used is composed of 108 examples, including 32 grammatical examples and 76 ungrammatical examples. Grammatical examples were constructed by the author based on examples from Dryer 2007b. Phenomena covered include the three types varying in attributive word order, attributive adjective agreement, copula agreement, copula word order, and tense.

The analysis of French utilized defines these three core syntactic types (regular, BAGS, and figurative) with an additional morphological type. The syntactic types are defined with the proper word order and given the supertype of the morphological type. A position class is then defined taking the morphological type as input and adding the proper agreement morphology to the adjectives.

To analyze the copula types in French, I defined a supertype for each of present tense and past tense. Then, each of these has the appropriate number of subtypes that specify each orthographical stem.

### *Penobscot [aaq-pen]*

Penobscot, an endangered language spoken in Maine, USA, can be described as a polysynthetic language, where adjectives appear as incorporated stems on noun stems. The Penobscot test suite is composed of 15 test cases, 9 of which are grammatical. The data for this test suite was retrieved from Quinn 2006, along with negative examples I created. Because adjectives are morphological in Penobscot, there is less (for predicative adjectives) or no (for incorporated adjectives) variation

in word order, resulting in fewer ungrammatical examples as a result of word order. Adjectives appear attributively as incorporated stems and can also be predicative, in which they appear as stative predicates.

Importantly, under the existing analysis of the Grammar Matrix customization system (Saleem 2010), verbs and adjectives in Penobscot are marked with argument markers when arguments are not overt. The existing analysis for verbs of these optional arguments transferred properly from verbs to predicative adjectives.

Note also that the target of this grammar is a regularized representation of the language, simplifying the morphophonology. This is based on the morphemic segmentation found in Quinn 2006.

A sampling of the Penobscot test suite is presented below:

- (142) a. wətihlaw mətahsəmal  
           wə-ih-l-α-w mat-ahsəm-al  
           3-tell-NA\_O-DIR-3 bad-dog-OBV  
           ‘he tells the bad dog’ [aaq-pen] (Quinn 2006)
- b. mkaséwiko  
           məhkasew-k-i-w  
           black-COP-NA\_O-3  
           ‘3.SG.NA is black’ [aaq-pen] (Quinn 2006)

I constructed a choices file for this test suite which analyzes adjectives as appearing as both attributive and predicative, though the attributive use is less common than incorporated stems. I defined incorporated adjectives as an incorporated stem lexical rule type within a position class on nouns. Predicative adjectives have an obligatory affix in their predicative form.

### *Maori [mri]*

Maori is a Malayo-Polynesian language spoken by approximately 148,000 people, primarily in New Zealand.<sup>5</sup> Maori is an example of a switching language, where predicative adjectives appear either with nominal morphology (definiteness agreement) or verbal morphology (aspect marking) (Stassen

---

<sup>5</sup>Population estimate from Ethnologue: <http://www.ethnologue.com/language/mri>, accessed on 11/24/14

2013). In addition, Maori attributive adjectives also must be the only modifier of the noun they modify (Bauer et al. 1997). Bauer et al. 1997 describes various Maori lexical types as having several particles that mark features on nouns and agreement on adjectives. I consider these sorts of particles beyond the scope of my work and are not supported by the existing customization system, so I approximated these particles as affixes and modeled them with lexical rules. An example of the data follows:

- (143) a. ka oma te kootiro  
           ka=oma te=kootiro  
           INCEP=run DEF=girl  
           ‘the girl runs’ [mri] (Biggs 1969)
- b. ka pai te whare  
           ka=pai te=whare  
           INCEP=good DEF=house  
           ‘the house is good’ [mri] (Biggs 1969)
- c. he pai te koorero  
           he=pai te=koorero  
           INDEF=good DEF=talk  
           ‘the talk is good’ [mri] (Biggs 1969)
- d. \* pai te whare te  
           pai=te whare=te  
           good=DEF house=DEF  
           intended: ‘the house is good’ [mri]

The Maori test suite is comprised of 23 examples, 10 of which are grammatical.

### *German [deu]*

German is of interest for two reasons: first, German adjectives agree with the nouns they modify in number and gender, and are also marked for case; second, German adjectives agree with nouns they modify, but do not agree with their subjects.

These two features are modeled in the following way. One adjective lexical type is defined with underspecified syntactic behavior. An obligatory position class is defined taking this lexical type as its input. This position class has two sorts of lexical rule types. First is a lexical rule type that

enables the adjective to be predicative without any affixes. Second is a hierarchy of lexical rule types defining each of agreement affixes for attributive adjectives. The attributive word order (pre-head) and predicative behavior (copula complement) is defined on the appropriate lexical rules. With this switching position class, I am able to model a test suite of 17 strings, 5 of which are grammatical. For German, I used grammatical examples from Landman and Morzycki 2002 and Hankamer and Lee-Schoenfeld 2005, as well as ungrammatical examples I created.

### *Russian [rus]*

Russian predicative adjectives are optionally copula complement, appearing as stative predicates in the present tense and as copula complements in the past tense. Russian adjectives also agree with both their modificand and subject in gender and number, and are marked for case. Unlike German, where adjectives are not marked in the predicative form, Russian predicative adjectives are marked for nominative or instrumental case in predicative constructions, but for the purposes of this grammar, I assume they are nominative.

My analysis of Russian utilizes the lexical rules for optionally copula complement adjectives to get the tense values right for adjectives. Stative predicate adjectives in Russian are always present tense, so the stative predicate lexical rules are constrained to be present tense. In copula complement constructions, the copula is marked for tense.

The Russian test suite is comprised of 21 strings, 8 of which are grammatical. My analysis of Russian is able to correctly model 100% of this test suite. My analysis is also built on top of previous grammar engineering done by students as collected in Language CoLLAGE (Bender 2014), utilizing the test suite and basic choices files.<sup>6</sup> However, I did remake the lexicon due to the differences in lexical content between my test suite and the choices file.

### *Illustrative language results*

As detailed in Figure 6.10, my analysis and implementation are able to accurately capture the behavior of each of the languages analyzed during development. Note that French and Penobscot

---

<sup>6</sup>Language CoLLAGE resources for Russian (rus) were created by Varvara Gracheva and Esad Suskic in 2010. Additional judgments were provided by Olga Zamaraeva.

| language  | positives | negatives | parses | coverage (TP) | overgeneration (FP) | spurious ambiguity |
|-----------|-----------|-----------|--------|---------------|---------------------|--------------------|
| English   | 13        | 44        | 13     | 100%          | 0%                  | 0%                 |
| French    | 32        | 76        | 35     | 100%          | 0%                  | 0%                 |
| Penobscot | 9         | 7         | 12     | 100%          | 0%                  | 0%                 |
| Maori     | 10        | 13        | 10     | 100%          | 0%                  | 0%                 |
| German    | 5         | 12        | 5      | 100%          | 0%                  | 0%                 |
| Russian   | 8         | 13        | 8      | 100%          | 0%                  | 0%                 |

Figure 6.10: Illustrative language test suite overview and results

| language  | syntactic behavior             | modification direction      | copula?    | inflection                    | incorporation |
|-----------|--------------------------------|-----------------------------|------------|-------------------------------|---------------|
| English   | both, attributive, predicative | pre-head                    | obligatory | none                          |               |
| French    | both                           | pre-head, post-head, either | obligatory | gender, number                |               |
| Penobscot | both                           | pre-head                    | impossible | none                          | X             |
| Maori     | both                           | post-head                   | impossible | definiteness <b>or</b> aspect |               |
| German    | both                           | pre-head                    | obligatory | gender, number, case          |               |
| Russian   | both                           | pre-head                    | optional   | gender, number, case          |               |

Figure 6.11: Illustrative language choices

both have ambiguous sentences, resulting in additional parses. Figure 6.11 details how each of the features described and implemented is used in these analyses.<sup>7</sup>

### 6.2.2 Held Out Languages

The pseudo-language and illustrative language tests detailed in the previous sections provide verification that the system functions as I intend. However, it is also important to test the system on new languages not considered during development to ensure that the system generalizes to unseen data. Subsequently, I developed and tested five additional grammars for languages that were not consid-

<sup>7</sup>Test suites available at <http://www.trimbleworks.us/linguistics/> or as part of the SVN repository at <svn://lemur.ling.washington.edu/shared/matrix/trunk>. Instructions for downloading at <http://www.delph-in.net/matrix/>



Figure 6.12: General geographic distribution of held out languages

ered during development. While these languages were not considered during development, several of them were catalogued by typologists in sources I referenced. These languages were chosen due primarily to their geographic and genealogical variety.

1. Luo [luo] (Eastern Sudanic, Nilo-Saharan)
2. Yup'ik, Central Alaskan [esu] (Eskimo, EskimoAleut)
3. Lakota [lkt] (Western Siouan, Siouan)
4. Mandarin [cmn] (Sinitic, Sino-Tibetan)
5. Frisian [frr] (Germanic, Indo-European)

The geography of these is shown in Figure 6.12. This section will summarize the analysis of each of these in turn.

### *Luo [luo]*

Luo, also known as Dholuo, is a language spoken by about six million people in Kenya and Tanzania.<sup>8</sup> Luo was chosen as a test language because it shows behavior which can be analyzed as

---

<sup>8</sup>Population estimate from Ethnologue: <http://www.ethnologue.com/language/luo>, accessed on 11/24/14

optional inflection.

Luo adjectives optionally appear without overt subjects and are marked for agreement with the subject in person and number. Luo can be analyzed as a switching language in that Luo predicative adjectives have optional inflection, sometimes marked for subject agreement like Luo verbs and sometimes not marked. Tucker and Bryan 1966 describe this as a distinction of aspect. In contrast to this, Stafford 1967 describes similar Luo adjectival constructions as predicative and attributive constructions with the attributive construction affixed with an attributive marker. There also seems to be a class of adjectives which do not inflect, as evidenced in Stafford 1967.

I decided to model this distinction as between predicative and attributive adjectives, along the lines of Stafford 1967. To model this, an adjective lexical type is defined to model the stems along with a switching position class with several lexical rule types. These types are primarily divided by syntactic behavior: the attributive marker is modeled by a lexical rule type that can be attributive while the predicative agreement is modeled with several lexical rule types that have to appropriate agreement features and affixes.

Agreement in plural with the subject could be described as a circumfix. I utilize two position classes to model this, the one described above and an additional to mark singular or plural agreement, where the singular agreement is null and the plural agreement is marked. The subject affix is also constrained for number, and therefore the proper affixes unify in the proper settings. Because both position classes are obligatory, both the proper person and number constraints are placed on the external argument.

A sample of the Luo data follows:

- (144) a. alwongo  
           a-lwongo  
           1.SG-call  
           ‘I am calling’ [luo] (Tucker and Bryan 1966)
- b. aber  
           a-ber  
           1.SG-good.SG  
           ‘I am good’ [luo] (Tucker and Bryan 1966)
- c. nyathi matin chiemo

nyathi ma-tin chiemo  
 child small eat.PROG  
 ‘the small child is eating’ [luo] (Stafford 1967)

- d. \* ber  
     ber  
     good  
     intended: ‘I am good’ [luo]

The Luo test suite has 31 strings, 10 of which are grammatical.

It should be noted that an implementation error resulted in the initial grammar produced for Luo to be unusable in that the stative predicate lex rule type was assigned as a supertype without including the definition. This issue was easily solvable by removing this spurious supertype from *luo.tdl*.<sup>9</sup>

### *Yup’ik [esu]*

Central Alaskan Yup’ik is a language spoken by about 19,000 people in Alaska.<sup>10</sup> Yup’ik is a polysynthetic language with much affixation, noun incorporation, and adjective incorporation (Mithun 2009). The language was chosen for this test as a language outside of the Algic family with incorporated adjectives. According to Miyaoka (2012), Yup’ik adjectives generally come in two separate stems, incorporated and predicative. The incorporated adjectives appear as affixes to the noun they modify as one stem, while its semantically equivalent predicative form takes another stem. Yup’ik also has a complex morphophonology, so I targeted a regularized underlying form for my grammar as generally described by Miyaoka 2012. Some examples follow:

- (145) a. anguq  
           ange-uq  
           big-3.SG  
           ‘he/she/it is big’ [esu] (Jacobson 1985)
- b. augna-arnaq mikuq  
     augna-arnaq mike-uq

---

<sup>9</sup>This bug has since been resolved, and the Luo grammar is generated properly.

<sup>10</sup>Population estimate from: [http://www.census.gov/hhes/socdemo/language/data/acs/SupplementaryTable1\\\_ACSB10-10.xls](http://www.census.gov/hhes/socdemo/language/data/acs/SupplementaryTable1\_ACSB10-10.xls), accessed on 11/24/14



PROXIMAL-woman small-3.SG  
 ‘that woman is small’ [esu] (Jacobson 1985)

- c. tunellruat enpaarrluk  
 tunellruat ena-paarrluk  
 sell.3.PL.PST house-big  
 ‘they sold a big house’ [esu]

I analyze the proximal markings with lexical rules, though they could probably be analyzed as optional determiners. A lexical type is defined for predicative adjectives, which are stative predicates, while attributive adjectives are defined as incorporated stems on nouns. The Yup’ik test suite has 13 strings, 6 of which are grammatical. The analysis and implementation described here is able to properly analyze these sentences.

### *Lakota [lkt]*

Lakota is a language spoken by about 6,000 speakers in North and South Dakota, USA.<sup>11</sup> Lakota, similar to Yup’ik and Penobscot, does not have overt independent pronouns. However, unlike these languages, Lakotan adjectives appear as both stand alone attributive and predicative adjectives as opposed to incorporated adjectives. Attributive adjectives appear post-head, while predicative adjectives are stative predicates. These stative predicate adjectives are sometimes called stative verbs, since they have similar morphology to verbs (Ullrich 2011). The following is a sample of the test suite:

- (146) a. mní kiŋ šmé  
 mní kiŋ šma  
 water DET.DEF deep  
 ‘the water is deep’ [lkt] (Ullrich 2011)
- b. mniŋhúha ska waŋ ophéwathuŋ  
 mniŋhúha ska waŋ ophé-wa-thuŋ  
 cloth white DET.INDEF buy-1.PL.AGT  
 ‘we bought some white cloth’ [lkt] (Ullrich 2011)

---

<sup>11</sup>Population estimate from Ethnologue: <http://www.ethnologue.com/language/lkt>, accessed on 11/24/14

- c. \* ska mniǰúha waŋ ophéwathuŋ  
 ska mniǰúha waŋ ophé-wa-thuŋ  
 white cloth DET.INDEF buy-1.PL.AGT  
 intended: ‘we bought some white cloth’ [lkt]

The Lakota test suite consists of 70 items, 9 of which are grammatical. The test suite was constructed in reference to work included in Language CoLLAGE (Bender 2014).<sup>12</sup>

While my extensions to the Grammar Matrix produce a grammar of Lakota which produces a good analysis for all 9 grammatical items and correctly rules out the other 61 ungrammatical items, the analysis generates some spurious ambiguity related to optional arguments. Because adjectives can be both attributive and predicative and subjects can be dropped, the resulting grammar analyzes sentences with both an adjective and verb as a verb and a predicative adjective, both of which have dropped subjects. Additional work needs to be done to fine tune the interactions between dropped arguments and adjectives.

#### *Mandarin [cmn]*

Mandarin Chinese is estimated to be the language with the most fluent speakers in the world, with approximately one billion speakers.<sup>13</sup> Mandarin attributive adjectives are quite similar to their English counterparts, appearing pre-head with no inflection. However, Mandarin predicative adjectives are stative predicates. There is an optional marker on attributive adjectives, which I analyze as an affix. Additionally, the strings *na ben* and *zhe ben* are analyzed as demonstrative determiners, though I make no assertion of the scalability of this claim. The test suite was constructed in reference to work included in Language CoLLAGE (Bender 2014).<sup>14</sup> The following is some examples from the test suite:

- (147) a. wo yao naben xiao=de shu  
 wo3 yao4 na4ben3 xiao3=de shu1

---

<sup>12</sup>Language CoLLAGE resources for Lakota (lkt) were created by Chris Curtis and David McHugh in 2013.

<sup>13</sup>Population estimate from Ethnologue: <http://www.ethnologue.com/language/cmn>, accessed on 11/24/14

<sup>14</sup>The data utilizes Latin alphabet orthographic representations of the Mandarin character set with numeric indices to the particular Chinese character represented by homophonous strings. Language CoLLAGE resources for Mandarin (cmn) were created by David Bullock in 2006.

1.SG want DET.DISTAL small-ATTR book  
 ‘I want that small book’ [cmn]

b. zheben shu hao  
 zhe4ben3 shu1 hao3  
 DET.PROXIMAL book good  
 ‘This book is good’ [cmn]

c. \* wo yao naben shu xiao=de  
 wo3 yao4 na4ben3 shu1 xiao3=de  
 1.SG want DET.DISTAL book small-ATTR  
 intended: ‘I want that small book’ [cmn]

The Mandarin test suite consists of 22 items, 8 of which are grammatical. The analysis and implementation described here is able to properly analyze these sentences.

### *Frisian [frr]*

West Frisian is a language spoken by approximately 470,000 speakers in the northern regions of the Netherlands. Frisian adjectives appear pre-head attributively and in copula complement constructions predicatively. Adjectives do not agree with their subject or modificand, though the copula does agree with the subject. The following is an example from the test suite:

(148) a. di gurt wunderbar soldaat sleep  
 di gurt wunderbar soldaat sleep-t  
 DET.DEF big wonderful soldier sleep-3.SG  
 ‘the big wonderful soldier sleeps’ [frr]

b. hi es gurt  
 hi es gurt  
 3.SG COP.3.SG big  
 ‘he is big’ [frr]

c. \* di gurt slept soldaat  
 di gurt sleep-t soldaat  
 DET.DEF big sleep-3.SG soldier  
 intended: ‘the big soldier sleeps’ [frr]

| language | positives | negatives | parses | coverage (TP) | overgeneration (FP) | spurious ambiguity |
|----------|-----------|-----------|--------|---------------|---------------------|--------------------|
| Frisian  | 4         | 34        | 4      | 100%          | 0%                  | 0%                 |
| Luo      | 10        | 21        | 10     | 100%          | 0%                  | 0%                 |
| Yup'ik   | 6         | 7         | 6      | 100%          | 0%                  | 0%                 |
| Mandarin | 8         | 14        | 8      | 100%          | 0%                  | 0%                 |
| Lakota   | 9         | 61        | 12     | 100%          | 0%                  | 25%                |

Figure 6.13: Held-out language test suite overview and results

| language | syntactic behavior | modification direction | copula?    | inflection     | incorporation |
|----------|--------------------|------------------------|------------|----------------|---------------|
| Frisian  | both               | pre-head               | obligatory | none           |               |
| Luo      | both               | post-head              | impossible | person, number |               |
| Yup'ik   | predicative        | –                      | impossible | person, number | X             |
| Mandarin | both               | pre-head               | impossible | none           |               |
| Lakota   | both               | post-head              | impossible | none           |               |

Figure 6.14: Held-out language choices

The test suite contains 38 items, 4 of which are grammatical. The test suite was constructed in reference to work included in Language CoLLAGE (Bender 2014)<sup>15</sup>. The analysis and implementation described here is able to properly analyze these sentences.

### *Held-out Language Results*

As detailed in Figure 6.13,<sup>16</sup> my implementation produces analyses for 100% of held out language test adjective strings, while properly forbidding ungrammatical sentences. Note that there is some erroneous ambiguity in Lakota. Figure 6.14 details how each of the features described and implemented is used in these analyses. The held out languages provide a more even set of choices than the development languages, though all but Frisian lack a copula.

<sup>15</sup>Language CoLLAGE resources for Frisian (frr) were created by Nigel Kilmer and Woodley Packard in 2013.

<sup>16</sup>Spurious ambiguity was calculated as (number of unintended parses in test suite)/(number of parses in test suite)

### **6.3 Summary**

This chapter has detailed a two-pronged evaluation of the adjective extension presented in Chapter 4. First is the set of pseudo-languages which encode the range of phenomena I set out to implement (as discussed in §2.5). Second is a two-step process of evaluation over first a set of languages used to steer development of the system, and second a set of languages not considered during development to test the applicability to additional languages. Each test is a set of strings, grammaticality judgments, and semantic representations for each string, along with a choices file.

The evaluation described here shows the adjective library extensions to perform as expected, generating correct parses for 100% of the grammatical data while producing no parses for ungrammatical data. The system generates very little spurious ambiguity: in only one test language did the system produce unintended parses. However, this evaluation also uncovers areas of future research, such as an analysis for particles, clitics, and further work on the interaction between adjectives and other libraries of the customization system, mainly argument optionality (in §6.2.2).

## Chapter 7

**CONCLUSION & FUTURE WORK**

This thesis presents an extension to the Grammar Matrix to enable user-linguists to analyze adjectives in their target language. This implementation is built upon a survey of relevant typological literature, existing analyses of adjectives and copulas in other systems, and semantic and syntactic literature on adjectives and copulas. I propose a novel analysis of adjectives, relying in part on the functionality of the Grammar Matrix customization system, that utilizes both lexical and morphological constraints to capture the variability across and within languages, and present a set of user interface elements to succinctly define the behavior of adjectives in target languages. An evaluation is conducted on the implementation which shows that my extensions to the Grammar Matrix are capable of constructing grammars which produce analyses as intended for a wide variety of languages.

**7.1 Summary**

I set out this thesis with the following goals (in Chapter 1):

- Describe the range of variation of adjectives across the world's languages to define the scope of phenomena to be implemented.
- Develop HPSG analyses of the syntax and semantics of these phenomena.
- Encode the analyzed phenomena into set of user interface options for user-linguists to select in the Grammar Matrix customization system.
- Develop a set of tests to evaluate the analyses and prevent regressions in later system development.

Chapter 2 has provided a thorough overview of the typology of adjectives, including the morphology, syntax, and semantics of adjectives in many languages. Chapter 4 has provided HPSG analyses of the syntax, morphology, and semantics of these phenomena. §5.1 has provided a detailed description of the encoding of these analyses into user interface options, while §5.2 covers

how these options are utilized to calculate types for the output grammar. Chapter 6 has presented a set of tests which evaluate my work and provide a suite of regression tests. The work presented here achieves 100% coverage with little spurious ambiguity over test suites for 11 natural languages.

## 7.2 Future Work

Along with the novel analyses I present, I also have discovered several areas ripe for future work. These can be divided into two groups: first, work on additional adjective and non-verbal predicate constructions; second, work on further incorporating adjectives into other libraries of the Grammar Matrix.

There is much additional work required to capture the range of adjective constructions in the world's languages. First, the analysis I present is concerned wholly with the semantics of intersective adjectives. An HPSG analysis of the semantics of subsective and non-intersective adjectives is warranted. Second, adjectival constructions appearing without a noun, such as *I'll take the red*, present a whole range of variations across languages that I did not consider in depth. Third, an analysis of comparative and superlative adjectival constructions could be reasonably built into the customization system. Fourth, significant typological work is needed on the semantic and syntactic properties of scopal adjectives and how they might tie into a larger picture of adjectives cross-linguistically. Lastly, there is the phenomenon of transitivity in adjectives, such as those in *it is worth my time* or *I am short two pennies*.

Second, I have detailed a trove of interactions between adjectives and other phenomena. The first interaction and foremost to the development of the Grammar Matrix is treatment of non-matrix clauses, and how adjectives, especially predicative adjectives, behave differently in matrix clauses vs. non-matrix clauses. The second interaction is between negation, adjectives, and copulas, where there is significant study to be warranted. The third interaction is between adjectives and argument optionality, and how the system might be improved to properly capture the linguistic facts (see §6.2.2). See §2.2.1 for more on these intersections.

Additionally, there is further work to be done on agreement and morphology. While the existing morphotactics system of the Grammar Matrix is capable of handling the agreement and inflectional aspects of adjectives through morphology, there are many languages which utilize clitics or particles

to mark agreement or other semantic features on adjectives, phenomena not currently supported by the customization system. At the other end of the scale, there is much work to be done on improving the coverage of polysynthetic languages, including phenomena such as noun and adverb incorporation, and how these relate to adjectival incorporation.

### **7.3 Closing**

In closing, I have presented a broad typological and theoretical survey of adjectives, a typologically grounded, thorough theoretical analysis of adjectives and copulas, a user interface for developing language-specific analyses, and an implementation of this analysis for extending starter grammars into two new lexical categories. I have presented a precise analysis of a large range of phenomena and an implementation that demonstrates their effectiveness. I hope future grammarians will find my extensions insightful and utilize them to capture linguistic facts of yet to be analyzed languages. The system is freely available for use online.<sup>1</sup>

---

<sup>1</sup>To use an up to date, active version of the Grammar Matrix, see: <http://www.delph-in.net/matrix/customize/matrix.cgi>. A frozen version of my instance is available at <http://trimbleworks.us/matrix/frozen/>. For a copy of the source files for my evaluation instance, go to <http://trimbleworks.us/linguistics/>.



## **Bibliography**

- Abraham, P. T. 1985. *Apatani Grammar*, volume 12 of *CIIL Grammar Series*. Manasagangotri, Mysore: Central Institute of Indian Languages.
- Alexiadou, Artemis and Wilder, Chris. 1998. Adjectival modification and multiple determiners. *Possessors, Predicates and Movement in the DP* pages 303–332.
- Andrade, Manuel J. 1919. The distinction between *ser* and *estar*. *Hispania* 2(1), 19–23.
- Attia, Mohammed, Butt, Miriam and King, Tracy Holloway. 2008. A unified analysis of copula constructions in LFG. In *Proceedings of the LFG08 Conference, CSLI Publications*, volume 3, page 105.
- Baker, Mark C. 1985. *Incorporation, a theory of grammatical function changing*. Doctoral Dissertation, Massachusetts Institute of Technology.
- Baker, Mark C. 2003. *Lexical categories: Verbs, nouns and adjectives*, volume 102. Cambridge University Press.
- Bauer, Winifred, Parker, William et al. 1997. *The Reed reference grammar of Māori*. New Zealand: Reed Publishing.
- Bender, Emily M. 2007. Combining Research and Pedagogy in the Development of a Crosslinguistic Grammar Resource. In Tracy Holloway King and Emily M. Bender (eds.), *Proceedings of the GEAF07 Workshop*, pages 26–45, Stanford, CA: CSLI.
- Bender, Emily M. 2008. Radical non-configurationality without shuffle operators: An analysis of Wambaya. In *Proceedings of the 15th International Conference on Head-Driven Phrase Structure Grammar, NICT, Keihanna, Japan*, pages 6–24.
- Bender, Emily M. 2014. Language CoLLAGE: Grammatical Description with the LinGO Grammar Matrix. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk and Stelios Piperidis (eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 2447–2451, Reykjavik, Iceland: European Language Resources Association (ELRA).

- Bender, Emily M., Drellishak, Scott, Fokkens, Antske, Poulson, Laurie and Saleem, Safiyyah. 2010. Grammar Customization. *Research on Language & Computation* 8(1), 23–72.
- Bender, Emily M., Flickinger, Dan and Oepen, Stephan. 2002. The Grammar Matrix: An Open-Source Starter-Kit for the Rapid Development of Cross-Linguistically Consistent Broad-Coverage Precision Grammars. In *Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics*, pages 8–14, Taipei, Taiwan.
- Bender, Emily M and Good, Jeff. 2005. Implementation for discovery: A bipartite lexicon to support morphological and syntactic analysis. In *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, volume 41, pages 1–16, Chic Ling Society.
- Bhat, D.N.S. 1994. The adjectival category: criteria for differentiation and identification. *Studies in the Language Companion Series* 24 page 295.
- Biggs, Bruce. 1969. *Let's Learn Maori*. Wellington: Reed.
- Bleam, Tonia M. 2000. Leísta Spanish and the syntax of clitic doubling. *IRCS Technical Reports Series* page 33.
- Bolinger, Dwight L. 1944. More on Ser and Estar. *The Modern Language Journal* 28(3), 233–238.
- Bolinger, Dwight L. 1947. Still more on ser and estar. *Hispania* pages 361–367.
- Bresnan, Joan and Mchombo, Sam A. 1995. The lexical integrity principle: Evidence from Bantu. *Natural Language & Linguistic Theory* 13(2), 181–254.
- Carlson, Greg N. 1977a. A unified analysis of the English bare plural. *Linguistics and philosophy* 1(3), 413–457.
- Carlson, Gregory N. 1977b. *Reference to kinds in English*. Doctoral Dissertation, University of Massachusetts, Amherst.
- Carnie, Andrew Hay. 1995. *Non-verbal predication and head-movement*. Doctoral Dissertation, Massachusetts Institute of Technology.

- Cinque, Guglielmo. 2010. *The syntax of adjectives: A comparative study*, volume 57. MIT Press.
- Copestake, Ann. 2002. Definitions of Typed Feature Structures. In Stephan Oepen, Dan Flickinger, Jun-ichi Tsujii and Hans Uszkoreit (eds.), *Collaborative Language Engineering*, pages 227–230, Stanford, CA: CSLI Publications.
- Copestake, Ann, Flickinger, Dan, Pollard, Carl and Sag, Ivan A. 2005. Minimal Recursion Semantics: An Introduction. *Research on Language & Computation* 3(4), 281–332.
- Copestake, Ann A and Flickinger, Dan. 2000. An Open Source Grammar Development Environment and Broad-coverage English Grammar Using HPSG. In *LREC*.
- Crespo, Luis. 1946. Los verbos SER y ESTAR explicados por un nativo. *Hispania* 29(1), 45–55.
- Crowell, Thomas Harris. 1979. *A grammar of Bororo*. Doctoral Dissertation, Cornell University.
- Crowgey, Joshua. 2012. *The Syntactic Exponence of Negation: A Model for the LinGO Grammar Matrix*. M.A. Thesis, University of Washington.
- Crowgey, Joshua. 2014. Constructing a morphophonological analyzer for Lushootseed. In Sadlier-Brown Emily Weber, Natalie and Erin Guntly (eds.), *UBC Working Papers in Linguistics*, volume 37, pages 39–63.
- Dalrymple, Mary, Dyvik, Helge and King, Tracy Holloway. 2004. Copular complements: closed or open. In *Proceedings of the LFG04 Conference*, pages 188–198.
- Dixon, Robert M. W. 1977. *A Grammar of Yidin*, volume 19 of *Cambridge Studies in Linguistics*. Cambridge: Cambridge University Press.
- Dixon, Robert M. W. 1982. ‘Where Have All the Adjectives Gone?’ and other Essays in Semantics and Syntax. Berlin: Mouton de Gruyter.
- Dixon, Robert M. W. 2004. Adjective Classes in Typological Perspective. In *Adjective Classes: A Cross-linguistic Typology*, pages 1–49, London: Oxford University Press.
- Drellishak, Scott. 2009. *Widespread but Not Universal: Improving the Typological Coverage of the Grammar Matrix*. Doctoral Dissertation, University of Washington.

- Dryer, Matthew. 2007a. Noun phrase structure. *Complex Constructions, Language Typology and Syntactic Description*, Vol. 2 2, 151–205.
- Dryer, Matthew. 2007b. Word order. *Clause Structure, Language Typology and Syntactic Description*, Vol. 1 1, 61–131.
- Dryer, Matthew S. 2013. *Order of Adjective and Noun*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Eades, Diana. 1979. Gumbaynggir. *Handbook of Australian languages* 1, 244–361.
- Emenanjo, Nolue. 1978. *Elements of Modern Igbo Grammar - a descriptive approach*. Ibadan, Nigeria: Oxford University Press.
- Engelberg, Stefan. 2005. Kimian states and the grammar of predicative adjectives. *Theoretical linguistics* 31(3), 331–347.
- Falk, Yehuda. 2006. *Subjects and universal grammar: An explanatory theory*, No. 113, Cambridge University Press.
- Flickinger, Dan. 2000. On Building a More Efficient Grammar by Exploiting Types. *Natural Language Engineering* 6 (1), 15–28.
- Flickinger, Dan, Bender, Emily M and Oepen, Stephan. 2003. MRS in the LinGO Grammar Matrix: A practical users guide. Technical Report, Technical report, Deep Thought Project Deliverable 3.5.
- Fokkens, Antske. 2011. Metagrammar Engineering: Towards Systematic Exploration of Implemented Grammars. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, HLT '11, pages 1066–1076, Stroudsburg, PA, USA: Association for Computational Linguistics.
- Fokkens, Antske. to appear. *Enhancing Empirical Research for Implemented Grammars*. Doctoral Dissertation, University of Saarland.
- Frajzyngier, Zygmunt. 1993. *A Grammar of Mupun*. Berlin: Dietrich Reimer Verlag.

- Gallego, Ángel J. 2006. A note on *lo que*, unpublished, available at [http://filcat.uab.cat/clt/membres/professors/gallego/pdf/Lo\\_que.pdf](http://filcat.uab.cat/clt/membres/professors/gallego/pdf/Lo_que.pdf).
- Goodman, Michael Wayne. 2013. Generation of machine-readable morphological rules from human-readable input. *Seattle: University of Washington Working Papers in Linguistics* 30.
- Greene, David. 1966. *The Irish Language*. Dublin: The Three Candles.
- Hahm, Hyun-Jong, Wechsler, Stephen, Butt, Miriam and King, Tracy Holloway. 2007. Untangling the Russian predicate agreement knot. In *The Proceedings of the LFG07 Conference*. Stanford CA: CSLI Publications, pages 233–249.
- Haiman, John. 1978. Conditionals are Topics. *Language* 54, 564–589.
- Hankamer, Jorge and Lee-Schoenfeld, Vera. 2005. What Moves in German VP-Fronting? Handout of presentation at Berkeley Syntax/Semantics Circle.
- Harms, Philip Lee. 1994. *Epena Pedee Syntax*, volume 4 of *Studies in the Languages of Colombia*. Dallas / Arlington: Summer Institute of Linguistics and University of Texas.
- Jacobson, Steven A. 1985. Siberian Yupik and Central Yupik Prosody. In M. Krauss (ed.), *Yupik Eskimo Prosodic Systems: Descriptive and Comparative Studies*, pages 25–46, Fairbanks: Alaska Native Language Center.
- Jary, Mark. 2004. Indicative mood, assertoric force and relevance. *Philosophica* 4, 2.
- Kang, Haeng Soon. 2005. On the adjective in Korean. *Rivista di Grammatica Generativa*, vol. 30 (2005), p. 3-15 .
- Kennedy, Christopher. 2012. Adjectives. In Delia Graff Fara and Gillian Russell (eds.), *Routledge Companion to Philosophy of Language*, pages 328–341, London: Routledge.
- Knop, Constance K. 1971. Mnemonic devices in teaching French. *French Review* 45, 337–342.
- Ko, Heejeong. 2005. *Syntactic edges and linearization*. Doctoral Dissertation, Massachusetts Institute of Technology.

- Laenzlinger, Christopher. 2005. French adjective ordering: Perspectives on DP-internal movement types. *Lingua* 115(5), 645–689.
- Landman, Meredith and Morzycki, Marcin. 2002. Event-kinds and the representation of manner. In *Proceedings of the Western Conference in Linguistics (WECOL)*, volume 11.
- Langdon, Margaret H. 1970. *A Grammar of Diegueo: The Mesa Grande Dialect*, volume 66 of *University of California Publications in Linguistics*. Berkeley: University of California Press.
- Larson, Richard. 1995. Olga is a beautiful dancer. In *Winter Meetings of the Linguistic Society of America, New Orleans*.
- Lee-Schoenfeld, Vera. 2007. *Beyond coherence: The syntax of opacity in German*, volume 114. John Benjamins Publishing.
- Li, Kening. 2008. Contrastive focus structure in Mandarin Chinese. In *Proceedings of the 20th North American Conference on Chinese Linguistics (NACCL-20)*, volume 2, pages 759–774.
- Maling, Joan. 1983. Transitive adjectives: A case of categorial reanalysis. In *Linguistic categories: Auxiliaries and related puzzles*, pages 253–289, New York: Springer.
- Mallinson, Graham and Blake, Barry J. 1981. *Language typology: Cross-linguistic studies in syntax*, volume 46. North-Holland Amsterdam.
- McFadden, Thomas. 2006. German inherent datives and argument structure. *Datives and Other Cases. Amsterdam: Benjamins* pages 49–77.
- Mithun, Marianne. 2009. Polysynthesis in the Arctic. *Variations on Polysynthesis: The Eskaleut Languages* 86, 3–17.
- Miyaoka, Osahito. 2012. *A Grammar of Central Alaskan Yupik (CAY)*, volume 58. Berlin: Walter de Gruyter.
- Nishiyama, Kunio. 1999. Adjectives and the copulas in Japanese. *Journal of East Asian Linguistics* 8(3), 183–222.

- O'Hara, Kelly. 2008. *A morphotactic infrastructure for a grammar customization system*. M.A. Thesis, University of Washington.
- Osborne, C. R. 1974. *The Tiwi Language*, volume 55 of *Australian Aboriginal Studies*. Canberra: Australian Institute of Aboriginal Studies.
- Partee, Barbara. 1995. Lexical semantics and compositionality. *An invitation to cognitive science: Language* 1, 311–360.
- Pollard, Carl and Sag, Ivan A. 1994. *Head-Driven Phrase Structure Grammar*. Chicago, IL: The University of Chicago Press.
- Pustet, Regina. 2003. *Copulas: Universals in the Categorization of the Lexicon: Universals in the Categorization of the Lexicon*. London: Oxford University Press.
- Quinn, Conor. 2006. *Referential-access dependency in Penobscot*. Doctoral Dissertation, Harvard University.
- Rankin, Jamie and Wells, Larry D. 2010. *Handbuch zur deutschen Grammatik*. Cengage Learning.
- Richter, Helmut. 2006. German Declension. <http://hhr-m.userweb.mwn.de/de-decl/>.
- Saito, Mamoru. 1985. *Some asymmetries in Japanese and their theoretical implications*. Doctoral Dissertation, Massachusetts Institute of Technology.
- Saleem, Safiyyah. 2010. *Argument Optionality: A New Library for the Grammar Matrix Customization System*. M.A. Thesis, University of Washington.
- Schachter, Paul. 1987. Tagalog. In Bernard Comrie (ed.), *The Worlds Major Languages*, pages 936–958, London: Oxford University Press.
- Schachter, Paul and Shopen, Timothy. 2007. Parts-of-speech Systems. *Complex Constructions, Language Typology and Syntactic Description, Vol. 1* pages 1–60.
- Siegel, Melanie and Bender, Emily M. 2002. Efficient Deep Processing of Japanese. In *Proceedings of the 3rd Workshop on Asian Language Resources and International Standardization at the 19th International Conference on Computational Linguistics*, Taipei, Taiwan.

- Song, Sanghoun. 2013. *A Grammar Library for Information Structure*. Doctoral Dissertation, University of Washington.
- Sproat, Richard and Shih, Chilin. 1991. The Cross-linguistic Distribution of Adjective Ordering Restrictions. In *Interdisciplinary Approaches to Language*, pages 565–593, New York: Springer.
- Stafford, R. L. 1967. *An Elementary Luo Grammar, with vocabularies*. Nairobi: Oxford University Press.
- Stassen, Leon. 2003. *Intransitive predication*. London: Oxford University Press.
- Stassen, Leon. 2013. *Predicative Adjectives*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Tucker, A. N. and Bryan, M. A. 1966. *Linguistic Analyses: the Non-Bantu Languages of North-Eastern Africa*. London: Oxford University Press.
- Ullrich, Jan. 2011. *New Lakota Dictionary*. Bloomington, IN: Lakota Language Consortium, second edition.
- Vendler, Zeno. 1968. *Adjectives and nominalizations*, No. 5, Berlin: Walter De Gruyter.
- Waugh, Linda R. 1977. *A semantic analysis of word order: Position of the Adjective in French*, volume 1. Brill Archive.
- Wetzer, Harrie. 1992. “Nouny” and “Verby” Adjectivals: a Typology of Predicate Adjectival Constructions, volume 10, pages 223–262. Berlin: Walter de Gruyter.
- Wetzer, Harrie. 1996. *The typology of adjectival predication*. Mouton de Gruyter.
- Zeller, Jochen. 2003. Word-level and phrase-level prefixes in Zulu. *Acta Linguistica Hungarica* 50(1), 227–244.