

©Copyright 2026

Emily Luedke

Next, Incorporate the Flour: A Recipe for Modeling Noun  
Incorporation in the LinGO Grammar Matrix

Emily Luedke

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

Master of Science

University of Washington

2026

Committee:

Emily M. Bender

Sanghoun Song

Program Authorized to Offer Degree:

Linguistics

University of Washington

## **Abstract**

Next, Incorporate the Flour: A Recipe for Modeling Noun Incorporation in the LinGO Grammar Matrix

Emily Luedke

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

This thesis describes the creation of a new library for modeling noun incorporation (NI) in the LinGO Grammar Matrix customization system. NI is the morphological attestation of a noun on a verb stem where the noun bears some semantic relation to the verb. Languages vary with respect to how NI functions. In some languages, incorporating verbs pattern like intransitives whereas in others, they pattern like transitives. Some languages allow other arguments to be promoted to the argument position vacated by the incorporated noun (IN). Some allow external elements to modify the IN. I present an analysis of the variation of NI within the syntactic framework of HPSG and semantic framework of MRS. I also describe how I implemented this analysis into the Grammar Matrix customization system to be used by user-linguists hoping to model NI in an implemented grammar. I evaluate my system on five illustrative languages (Chukchi [ckt], Southern Tiwa [tix], Mapudungun [arn], Mohawk [moh], and Tongan [ton]), a set of constructed pseudo-languages, and five held-out languages (Apuinã [apu], Bribri [bzd], Inuktitut [iku], Moloko [mlw], and Yaqui [yaq]) The results show that my library effectively generalizes to unseen data with an average of 95.6% coverage, 15.2% overgeneration, and an ambiguity of 1.7 parses per grammatical sentence averaged across each grammar.

## TABLE OF CONTENTS

	Page
List of Figures . . . . .	iii
Chapter 1: Introduction . . . . .	1
Chapter 2: Background . . . . .	3
2.1 Head-Driven Phrase Structure Grammar . . . . .	3
2.2 Minimal Recursion Semantics . . . . .	9
2.3 The Grammar Matrix and Customization System . . . . .	11
2.4 Summary . . . . .	17
Chapter 3: Literature Review . . . . .	18
3.1 Terminology . . . . .	18
3.2 Typology of Noun Incorporation . . . . .	20
3.3 Syntactic and Lexicalist Analyses of Noun Incorporation . . . . .	29
3.4 Classifying Types of Noun Incorporation . . . . .	29
3.5 HPSG and Noun Incorporation . . . . .	32
3.6 Summary . . . . .	47
Chapter 4: Analysis . . . . .	48
4.1 Worked Examples of Noun Incorporation . . . . .	49
4.2 Lexical and Phrase Structure Rules of Noun Incorporation . . . . .	59
4.3 Summary . . . . .	74
Chapter 5: Implementation . . . . .	76
5.1 Background . . . . .	76
5.2 Questionnaire Design . . . . .	77
5.3 Python Code Implementation . . . . .	86

5.4	Validation . . . . .	99
5.5	Summary . . . . .	104
Chapter 6:	Evaluation . . . . .	105
6.1	Illustrative Languages . . . . .	106
6.2	Pseudo-languages . . . . .	121
6.3	Held-out Languages . . . . .	128
6.4	Summary . . . . .	146
Chapter 7:	Conclusion . . . . .	148
7.1	Overview . . . . .	148
7.2	Future Work . . . . .	149
7.3	Summary . . . . .	150

## LIST OF FIGURES

Figure Number	Page
3.1 Noun Incorporation type constraint from Runner and Aranovich, 2003:366 . . .	34
3.2 Classifier NI type constraint from Runner and Aranovich, 2003:366 . . . . .	35
3.3 Compounding NI type constraint from Runner and Aranovich, 2003:368 . . .	36
3.4 Noun Incorporation lexical rule from Malouf, 1999:9 . . . . .	38
3.5 A lexical entry for <i>ammassak</i> ‘sardine’ . . . . .	39
3.6 A lexical entry for <i>-tor</i> ‘-eat’ . . . . .	39
3.7 The denominal verb <i>ammassattor</i> ‘sardine-eat’ . . . . .	40
3.8 Adjunct lexical rule from (Malouf, 1999:11) . . . . .	40
3.9 Structure of sentence (41) . . . . .	42
3.10 Noun Incorporation lexical rule from Ball, 2005:19 . . . . .	44
5.1 First checkbox on NI subpage indicating that NI occurs in the language . . .	78
5.2 Choices capturing the behavior of Reduction NI . . . . .	79
5.3 Example user input for a language where Reduction NI results in an intransi- tive sentence with a different case value on the subject. . . . .	80
5.4 Choices capturing the behavior of Promotion NI . . . . .	81
5.5 Choices capturing the presence or absence of Doubling and Stranding NI . .	82
5.6 Once NI is defined in the language, the user may now define Incorporated Stem Lexical Rule Types. . . . .	84
5.7 IS-LRT with features defined on the IN, orthography and predicate value defined.	85
5.8 A validation error showing that the user has not finished describing the be- havior of Reduction NI in the language . . . . .	101
5.9 A validation warning showing that a nested choice was made without choosing the corresponding checkbox . . . . .	103
6.1 Parse tree for sentence (108) . . . . .	119
6.2 Parse tree for sentence (110) . . . . .	120
6.3 Semantic representation for pseudo-language sentence (112) . . . . .	127

6.4	Two parses generated for sentence (118)	131
6.5	Semantic representation for sentence (130)	137
6.6	Semantic representation for sentence (131b)	139
6.7	Semantic representation for sentence (135)	142

## ACKNOWLEDGMENTS

I would first like to thank my advisor, Emily M. Bender, for all of her guidance, encouragement, and mentorship throughout my time in CLMS, and in particular during my work on this thesis. When I first approached her with the idea of building a library for noun incorporation, I wasn't sure if this was something I was capable of doing. Thanks to her, I grew to truly enjoy the process of writing, communicating my ideas more clearly with every mark of the (metaphorical) red pen.

I am also grateful to my reader, Sanghoun Song, for his insightful feedback. He has always made me feel welcome in this research community.

I want to thank the DELPH-IN members for welcoming me into their ranks and steering this project in the right direction.

I would also like to thank my cohort of CLMS students who kept me accountable during this process and who were always willing to lend an ear. I would not have gotten through this thesis without their support and friendship.

Finally, I want to thank my family. I am forever grateful to my mom and dad who encouraged me to pursue my interests wherever they may lead—whether across the world or across the country. And to my partner, Ryan. None of this would have been possible without you. Thank you for believing in me always.

## Chapter 1

# INTRODUCTION

This thesis describes the creation of a noun incorporation (NI) library for the LinGO Grammar Matrix customization system (Bender et al., 2002; Bender et al., 2010; Zamaeva et al., 2022). The Grammar Matrix is a project for jump-starting the development of implemented grammars, using the syntactic framework of Head-Driven Phrase Structure (HPSG) (Pollard and Sag, 1994) and semantic framework of Minimal Recursion Semantics (MRS) (Copestake et al., 2005). The customization system is made up of libraries which store analyses of various phenomena that can be useful to users, and outputs analyses consistent with given specifications. These specifications are elicited from an online questionnaire designed by authors of the libraries. This project makes the process of creating a precision grammar easier since linguists do not need to re-analyze the phenomena from scratch for each language.

The library presented here models NI cross-linguistically. NI is the morphological attestation of a noun on a verb stem where the noun bears some semantic relation to the verb. I provide a lexical analysis of a range of NI manifestations with the aim of generating semantic representations mimicking those of unincorporated paraphrases. NI ranges in its variation of expression. In some languages, the object of the verb incorporates and there is no overt expression of an object in the syntax. I refer to these kinds of constructions as Reduction NI. Reduction NI can result in a sentence which appears intransitive (i.e. in terms of agreement morphology or case markers), and other times it can appear transitive (except for the lack of syntactic object). Some languages allow other arguments to be promoted to object position when a noun is incorporated. I refer to these kinds of constructions as Promotion

NI. The promoted argument can either have a possessive relationship with the IN, or it can be an oblique argument. Some languages allow external phrases to modify an IN. I refer to these constructions as Modifier NI. External modifiers can be adjectives or nouns which are co-referential and more (or as) specific as the IN. I analyze these NI types independently, and allow users to include whichever combination best suits the language they are creating a grammar for.

This thesis is organized as follows: In chapter 2 I provide background information on the syntactic and semantic frameworks used in my analysis as well as a conceptual introduction into the larger project which this thesis contributes to. In chapter 3, I review the literature on NI, including a descriptive typological survey, previous analyses, and existing HPSG accounts of NI. Chapter 4 presents my own analysis of NI with descriptions of the desired semantic representations for various types, and the lexical and phrase structure rules needed to obtain them. In chapter 5 I show how I implemented this analysis into the Grammar Matrix customization system, including my additions to the user-facing webpage. Chapter 6 describes the results of my evaluation of my system using a series of development and test grammars to measure coverage, overgeneration, and ambiguity. Finally, I conclude and offer thoughts for future work on this library in chapter 7.

## Chapter 2

# BACKGROUND

In this chapter, I present an overview of the theoretical background for this thesis. I introduce the syntactic and semantic frameworks which I use when developing my analysis of noun incorporation. I also introduce the larger project which my thesis contributes to: The Grammar Matrix customization system.

### 2.1 *Head-Driven Phrase Structure Grammar*

Head-Driven Phrase Structure Grammar (HPSG) is a highly lexical, phrase structure grammar (Pollard and Sag, 1994). In HPSG, the lexicon plays a major part in the grammar, where lexical entries are richly encoded with constraints. HPSG maps features with values, and visually these mappings are represented in attribute-value matrices (AVMs), such as the one below.

$$(1) \left[ \text{PER} \quad 3rd \right]$$

This AVM shows a feature structure mapping the feature PER (for person) with the value *3rd* (for third). The value of a feature can be a simple string or boolean value, or it can be another feature structure.

All elements of a language's grammar from lexicon to syntax are encoded in feature structures. These include lexical entries, lexical rules, and phrase structure rules. Feature structures also encode semantic representations, which will be discussed in section 2.2.

Feature structures are typed. Types exist in a hierarchy, with those lower in the hierarchy inheriting all of the constraints of those higher. Features structures in a grammar will have a type, and may further encode new constraints that are not shared by other feature structures in the grammar. Types allow for words or rules which share characteristics to inherit these

characteristics from a shared supertype (one that both entities inherit from) rather than writing large, redundant, and unnecessary feature structure definitions into the grammar.

For instance, a grammar might have a type *verb-lex*, which intransitive and transitive verbs both inherit from. A difference between intransitive and transitive verbs is their ability to combine syntactically with a direct object. The definition for *verb-lex* adapted from a small grammar of English is given in (2).<sup>1</sup>

$$(2) \left[ \begin{array}{l} \textit{verb-lex} \\ \text{SYNSEM.LOCAL.CAT.VAL} \left[ \begin{array}{l} \text{SPEC} \langle \rangle \\ \text{SUBJ} \langle \boxed{1} \rangle \end{array} \right] \\ \text{ARG-ST} \langle \boxed{1}, \dots \rangle \end{array} \right]$$

This feature structure shows that verbs have an empty SPEC list, and that the sole element on their SUBJ list (the subject), is identified with the first element on the ARG-ST (argument structure) list. The definitions for intransitive and transitive verbs are shown in (3) and (4).

$$(3) \left[ \begin{array}{l} \textit{intransitive-verb-lex} \\ \text{SYNSEM.LOCAL.CAT.VAL.COMPS} \langle \rangle \end{array} \right]$$

$$(4) \left[ \begin{array}{l} \textit{transitive-verb-lex} \\ \text{SYNSEM.LOCAL.CAT.VAL.COMPS} \langle \boxed{1} \rangle \\ \text{ARG-ST} \langle \dots, \boxed{1} \rangle \end{array} \right]$$

Both of these types inherit from the *verb-lex* type. They also inherit any constraints that *verb-lex* does from its own supertypes. Intransitive verbs do not combine with direct objects, so they have an empty COMPS list. Transitive verbs do, so there is one element on their COMPS list, which is identified with the second element on the ARG-ST list. Since both kinds of verbs inherit from (*verb-lex*), the specification about the subject are applicable to both. A (more) fully described AVM for transitive verbs including constraints from *verb-lex* looks like (5).

---

<sup>1</sup>The choices file for this small grammar was provided as part of the University of Washington LING 567 course: <https://courses.washington.edu/ling567/>

$$(5) \left[ \begin{array}{l} \textit{transitive-verb-lex} \\ \text{SYNSEM.LOCAL.CAT.VAL} \\ \text{ARG-ST} \end{array} \left[ \begin{array}{ll} \text{SPEC} & \langle \rangle \\ \text{SUBJ} & \langle \text{①} \rangle \\ \text{COMPS} & \langle \text{②} \rangle \end{array} \right] \right]$$

Feature structures can inherit from multiple types, which is known as multiple inheritance, allowing the grammar to capture cross-cutting generalizations. For instance, *verb-lex* in the small English grammar inherits from multiple supertypes including *basic-verb-lex*, which defines the type of the verb’s semantic contribution, and *non-mod-lex-item*, which states that the verb is not a modifier of another element.

The feature structures for the verb types above are complex and merit further explanation. These types have two features: SYNSEM and ARG-ST. SYNSEM is a feature that takes another feature structure as a value (one feature of which is LOCAL,<sup>2</sup> which takes CAT, which takes VAL and so on). Its embedded feature structure values contain syntactic and semantic information. Syntactic information is stored in the feature CAT shown above. Any semantic content is stored in the CONT feature, which is not shown above. CAT contains both valence information (in VAL) and information about part of speech as well as containing information about what other kind of entity it may modify (in the MOD feature, also not shown) among other things (in HEAD). The features found in VAL represent the different syntactic arguments: subject (SUBJ), complement(s) (COMPS), and specifier (SPR). Another feature, SPEC, allows specifiers themselves to constrain values on the element they specify. Each of these valence features takes a list as a value, which may contain up to one feature structure representing an entity’s subject, specifier, or element specified, and multiple complements.

An empty list on a valence feature means that the element is not “searching” for an argument of that syntactic role; a non-empty list means that the element is. Feature structures

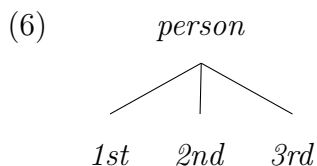
---

<sup>2</sup>LOCAL is a feature which contains syntactic and semantic information that is relevant in local contexts as opposed to the feature NONLOCAL, which contains information for non-local dependencies (Müller et al., 2024).

combine with other feature structures they are ‘searching’ for with phrase structure rules.

As mentioned, lexical and phrase structure rules are also represented by typed feature structures. A lexical rule has two main features: DTR and SYNSEM. DTR (pronounced daughter) represents the input to the lexical rule. SYNSEM is then the syntactic and semantic information about the output, also known as the mother of the rule. A lexical rule can specify values on either the daughter or the mother. It can also specify that the value is the same on the mother as it is on the daughter. This identification of values between mother and daughter happens via tagging. A value can be marked in the feature structure with a boxed tag (such as  $\boxed{\text{PER}}$  in example (5) above). Anywhere that tag appears in the AVM will have identical values.

The mechanics behind how lexemes are inflected and phrases are built in HPSG relies on the idea of unification. For two feature structures to unify, they must either be identical, or for each feature in the feature structure, one value must be less constrained (less specific but not contradictory). For example, let’s assume that in the grammar, there are three grammatical person values *1st*, *2nd*, and *3rd*. Each of these is of type *person*. This type hierarchy can be visualized in (6).



The two feature structures in (7) and (8) below would then unify, because the value of PER on the second is underspecified but still compatible with the fully specified value of the first.

(7)  $\left[ \text{PER } 3rd \right]$

(8)  $\left[ \text{PER } person \right]$

The feature in (9) structure would not unify with the one in (7), as its value is not identical to or less specific than the first. In other words, their values are conflicting.

$$(9) \left[ \text{PER } 2nd \right]$$

Unification is also used in phrase structure rules to combine words together. For example, in the noun phrase *the cookie*, there is a feature structure for each of the words, built from their lexical entries and inheriting constraints from any defined supertypes. Two such feature structures are shown in (10) and (11).

$$(10) \left[ \begin{array}{l} \text{STEM} \quad \langle \text{the} \rangle \\ \text{SYNSEM.LOCAL.CAT} \quad \left[ \text{HEAD } det \right] \end{array} \right]$$

$$(11) \left[ \begin{array}{l} \text{STEM} \quad \langle \text{cookie} \rangle \\ \text{SYNSEM.LOCAL} \quad \left[ \begin{array}{l} \text{CONT.HOOK.INDEX.PNG} \quad \left[ \begin{array}{l} \text{PER } 3rd \\ \text{NUM } sg \end{array} \right] \\ \text{CAT} \quad \left[ \begin{array}{l} \text{HEAD } noun \\ \text{VAL.SPR} \quad \langle \boxed{1} \left[ \text{LOCAL.CAT.HEAD } det \right] \rangle \end{array} \right] \end{array} \right] \\ \text{ARG-ST} \quad \langle \boxed{1} \rangle \end{array} \right]$$

The feature structure for *cookie* has the values *noun* for HEAD and *3rd* for PER, and specifies that there be an element on its SPR list with a HEAD value of *det* (for determiner). The determiner on the SPR list is identified with the sole element on the noun's ARG-ST list ( $\boxed{1}$ ). The feature structure for *the* says its HEAD value is *det*.

The phrase structure rule **spec-head-phrase** is given in (12).

$$(12) \left[ \begin{array}{l} \text{SYNSEM.LOCAT.CAT.VAL} \quad \left[ \begin{array}{l} \text{SPR} \quad \langle \rangle \\ \text{SPEC} \quad \boxed{1} \end{array} \right] \\ \text{HEAD-DTR.SYNSEM.LOCAL} \quad \left[ \begin{array}{l} \text{CAT} \quad \left[ \begin{array}{l} \text{HEAD} \quad \boxed{2} \\ \text{VAL} \quad \left[ \begin{array}{l} \text{SPR} \quad \boxed{3} \\ \text{SPEC} \quad \boxed{1} \end{array} \right] \end{array} \right] \\ \text{CONT.HOOK} \quad \boxed{4} \end{array} \right] \\ \text{NON-HEAD-DTR.SYNSEM} \quad \boxed{3} \quad \left[ \text{LOCAL.CAT.VAL.SPEC} \quad \left\langle \left[ \begin{array}{l} \dots \text{ HEAD} \quad \boxed{2} \\ \dots \text{ HOOK} \quad \boxed{4} \end{array} \right] \right\rangle \right] \end{array} \right]$$

This rule states that the specifier of the head daughter, in this case *cookie*, needs to unify with the entire feature structure for the non-head daughter, *the* (3). The only specification shown for *cookie*'s SPR value is that the HEAD be *det*, which is compatible, and indeed identical, with the head value of *the*. This rule also contains restrictions made by the specifier for this element it specifies. The element on the non-head daughter's SPEC list must have the same HEAD value (2) and HOOK value (4) as the head daughter. The lexical entry for *the* doesn't put a HEAD value restriction on the element on its SPEC list, so the HEAD value *noun* on *cookie* unifies. Thus, a noun phrase can be built with these two words as input using the **spec-head-phrase** rule.

To be more precise, we need to account for number marking on the noun. Before going through the phrase structure rule, *cookie* went through a non-inflecting lexical rule which specified that the NUM value be *sg* (for singular). The lexical entry for *cookie* is underspecified for number, and it must go through either the singular lexical rule, or an inflecting plural lexical rule before being an input to **spec-head-phrase**.

Phrase structure rules can be branching (such as the rule above) or non-branching. A non-branching rule creates a phrase out of a single word or phrase. An instance of this is the **bare-np** rule. Above, a noun phrase was created out of a noun and a specifier. However, not all nouns need specifiers such as proper nouns or plural nouns in English as in the sentence *Paula likes pastries*. Both *Paula* and *pastries* here go through the **bare-np** rule to become phrases that no longer 'search' for a specifier.

HPSG is a lexicalist framework which means that it analyzes the formation of words separately from the formation of phrases. This is known as the Lexical Integrity Principle (Bresnan and Mchombo, 1995). From this framework follow two important constraints. First, the syntax cannot be responsible for word formation. Second, parts of words cannot play a role in syntactic operations. In terms of the types of rules available in HPSG, this means that the output of lexical rules can serve as the input to other lexical rules or phrase structure rules; but the output of phrase structure rules can only feed other phrase structure rules. In other words, a word must be fully inflected before it forms a phrase. The specifications

of both lexical rules and phrase structure rules represent predictions about grammaticality in a language. Phrases and sentences licensed by these rules—and only the phrases and sentences licensed by these rules—are predicted to be grammatical in the language being modeled.

### 2.1.1 Summary

In this section, I introduced several key concepts of HPSG, including the notions of typed feature structures which map features with values, and the hierarchy they exist in. I described both lexical and phrase structure rules which are used in the formation of words and phrases, respectively. In the following section, I will describe the semantic framework which is integrated with HPSG in the Grammar Matrix to form semantic representations of the syntactic structures.

## 2.2 Minimal Recursion Semantics

The semantic framework used in the Grammar Matrix (described in section 2.3) and this thesis is Minimal Recursion Semantics (MRS). MRS (Copestake et al., 2005) was developed to be able to model a semantic representation in a flat manner compatible with the typed feature structures of HPSG. MRSs for phrases and sentences are built from values found in the CONT features of elements that make up the phrase or sentence.

An MRS is a bag of Elementary Predications (EPs) which each represent a semantic entity. An EP will have a handle, a relation, and arguments. The EP’s handle, of type *handle*, is the value of the feature LBL. PRED stores the relation of the EP. In the Grammar Matrix, the relation for EPs introduced by lexical entries is a simple string with the convention of the semantic contribution followed by a letter denoting the part of speech followed by “rel” (ex: `_make_v_rel` for the verb *make*) (Flickinger et al., 2003). For EPs introduced by constructions, this value follows the convention of the relation’s sense followed by “rel” (ex: `poss_rel` for modeling possession) (Nielsen, 2018). The arguments of an EP, labeled ARG $n$  (where  $n$  is 0, 1, 2, etc.) can either be of type *handle*, *ref-ind*, or *event*. The ARG0 of the

EP represents the index of the entity itself. Elements that refer to entities (such as nouns) will introduce EPs whose ARG0 is of type *ref-ind*, with features of noun-like elements such as person (PER) and number (NUM). Elements that refer to events (such as verbs) will introduce EPs whose ARG0 is of type *event* which encodes verb-like attributes such as tense (TENSE) and aspect (ASPECT). The other ARGs point to either the handles or the ARG0s of other EPs.

The interpretation of what each ARG $n$  of an EP points to depends on the nature of that EP. For verbs, the ARG1 is usually the more agent-like argument. An EP for the transitive verb *bake* would have the ARG1 pointing to the ‘baker’ (the noun on the verb’s SUBJ list), and the ARG2 pointing to the ‘baked thing’ (the noun on the verb’s COMPS list). An intransitive verb’s EP would then not have the ARG2 feature, and only point to the ‘doer’ of the verb. In the case of adjectives, the ARG1 points to the ARG0 of the noun it modifies.

All *ref-inds* must be bound by a quantifier. In the case where a noun occurs with a specifier, as in the example *the cookie* above, the lexical entry for the word *the* would contribute the necessary quantifier relation. Specifier-less nouns going through the **bare-np** rule on the other hand must be quantified with a quantifier relation introduced via the phrase structure rule itself.

A quantifier relation, in addition to its LBL and ARG0 has two other features: RSTR (for restriction) and BODY. These two features take handle values. Because of the possible ambiguity with scopal relations, a quantifier does not point directly to the noun via the RSTR feature. Instead, it is said to be ‘*qeq*’ with the noun. *qeq*, or equal, modulo quantifiers, is a relation that allows for another scoping relationship to come between the quantifier and the noun. In other words, the values are equal unless another quantifier intervenes. It consists of two features, HARG for the higher argument (the quantifier’s RSTR value) and LARG for the lower (the noun’s LBL). This allows for some scope taking element (another quantifier) to intervene between the noun and the quantifier introduced by its determiner. *qeq* relations are recorded in a list in the HCONS (for Handle CONStraints) feature of an MRS.

The example below for the sentence *I bake the cookie*, shows the identification of the



et al., 2010; Zamaraeva et al., 2022) are tools that grammar engineers can use as a starting point for building broad-coverage, precision implemented grammars using the syntactic and semantic frameworks described above. Implemented grammars, such as those built with this technology, are used for parsing and generation tasks, as well as linguistic hypothesis testing. Building an implemented grammar from scratch is a tedious task, and there would be redundant steps across the creation of multiple grammars. The Grammar Matrix answers the problem of redundancy by defining types for lexical and phrase structure rules to inherit from, as well as defining the necessary scaffolding in the grammar files such as definitions for lists and AVMs. The type information provided by the Matrix is hypothesized to be relevant for implemented grammars of any language. However, the Matrix is not itself a complete grammar, and must be elaborated with additional types specific to a language. The customization system aids in this step by making use of stored cross-linguistic HPSG and MRS analyses of various linguistic phenomena. These analyses, the elicitation questions in the customization system, and the corresponding backend code are known as libraries. Existing libraries which are relevant to my research include morphotactics (O’Hara, 2008; Goodman, 2013), valence change (Curtis, 2018), argument optionality (Saleem and Bender, 2010), adnominal possession (Nielsen, 2018).

The customization system is accessed via a website that a user seeking an implemented grammar can visit. This webpage consists of various subpages, each targeting a different phenomenon. Users will answer a series of questions designed to elicit the specific way the language handles that grammar phenomenon. Questions come in different formats, with some allowing users to select pre-defined options from a list, and others allowing the user to define new elements themselves. For instance, on the word order page, the user will select one option from a finite list of different word orders. On the lexicon page, the user will define an open-ended number of stems for different parts of speech and include the predicate value contributed by the stem. On the morphology page, the user will define the morphemes present, specifying the potential feature values contributed by each.

After interacting with questions on a page of the customization system website, the user

should select the ‘save and stay’ button which saves the current choices. If the combination of choices is invalid or undesirable, a validation error or warning will appear with an informative message encouraging the user to make a different selection or to fix an input mistake. When desired, and given that all errors have been addressed, the user may download a zipped file containing all of the elements of their grammar needed to compile with and use DELPH-IN<sup>3</sup> software for parsing and generation.

Each choice made by the user while interacting with the customization system is recorded in a text file known as the grammar specification or choices file. Every line in this file consists of a key and a value, with some keys elaborated so that they are referencing different features of the same element. For example, the lines in (14) are taken from a choices file for Southern Tiwa [tix]<sup>4</sup> used in the development of my library for noun incorporation.

```
(14)  verb1_name=trans-with-ni
       verb1_valence=trans
           verb1_stem1_orth=mu
           verb1_stem1_pred=_see_v_rel
           verb1_stem2_orth=kar
           verb1_stem2_pred=_eat_v_rel
           verb1_stem3_orth=tuwi
           verb1_stem3_pred=_buy_v_rel
```

The string ‘verb1’ in each line means that each provides specification to the same category of verb — in this case a transitive verb. Here, I provided the string ‘trans-with-ni’ as the name for the verb (as this is a class of verbs that can incorporate nouns), and selected transitive valence from a finite list of options. The indented lines represent three different verb stems I defined, each a transitive verb denoted by the `verb1_` prefix. The orthography (`orth`) and predicate (`pred`) for each stem are grouped by the stem index (1-3) in the key.

---

<sup>3</sup><https://github.com/delph-in>

<sup>4</sup>Here and throughout this paper, I provide the ISO 369 codes for languages for clarification.

The choices file is read in by the customization system, and based on the combinations of keys and values present, the backend code writes the specific types and rules that the language is analyzed to have. The choices file can also be used as a checkpoint. When a grammar is produced, one of the files present is a copy of the choices file used to generate it. This choices file can be input into the customization system to re-populate the choices visible on the webpage.

The contents of the grammar folder consist of files and scripts necessary for compatibility with DELPH-IN software, a copy of the choices file, and several files with the `.tdl` extension including the Grammar Matrix core grammar and language-specific files. These files are written in Type Description Language (TDL) which is a formalism for representing typed feature structure grammars computationally (Copestake, 2002b). `matrix.tdl` is the core set of types, and this exact file (barring any updates due to the continuous development of matrix libraries) is present in all generated grammars. `my_lang.tdl`, where *my\_lang* is substituted with the name of the language, is the extension of types from `matrix.tdl` specific to the language. The types in this file will inherit from those in `matrix.tdl` with general constraints, but contain more specific features, inflectional flags, etc. `rules.tdl` is a file that instantiates phrase structure rules found in the language, such as the `spec-head-phrase` rule described above. Similarly, `irules.tdl` and `lrules.tdl` instantiate inflecting lexical rule instances and lexical rule types respectively.

Though handling different grammatical phenomena, the different Matrix libraries do have a fair amount of interaction. For instance, noun incorporation involves the affixation of a noun root on a verb root, and therefore necessarily interacts with the morphotactics library (O’Hara, 2008; Goodman, 2013). Regression testing (Bender et al., 2007) is used to ensure that new analyses do not break existing ones where they interact. The regression tests consist of choices files and test suites. When the regression tests are run, they generate grammars with the given choices files, parse sentences in the testsuite, and compare the results with gold standard parses stored with the testsuites. As I developed my own library, I ran the regression tests at various points in development to see if any changes or additions I made to

files in the code base caused a failure. Then, at the end of developing my library, I added the choices files, test sentences and gold standard parses from the grammars used for evaluation in chapter 6 to the suite of regression tests for future developers to test against.

Noun incorporation, by the definition introduced in chapter 3, is a morphological process. The implementation involves the use of the morphotactics library similarly to how the adjectives library interacts with it to implement incorporated adjectives (Trimble, 2014). As will be explored in the following chapter, noun incorporation can impact the case marking on core arguments. Case marking is supported by the case library (Drellishak, 2009). Additionally, part of my analysis of noun incorporation involves necessitating the presence or absence of some arguments which is handled by the argument optionality library (Saleem, 2010; Saleem and Bender, 2010). I will briefly describe each of these libraries below.

The morphotactics library was originally developed by O’Hara (2008) and later expanded by Goodman (2013). This library provides functionality for three connected elements. The first is a position class (PC). The PC is a slot in the morphological expression of a stem. For instance, in English, nouns have a suffix slot for marking number. This suffix slot would be a PC. Each PC can then contain one or more lexical rule types (LRTs). In the number example introduced above, we would have an LRT for singular number, and an LRT for plural. Each LRT then consists of one or more lexical rule instances (LRIs), which contribute the actual affix. In the *cookies* example, the LRT for plural number would have an LRI with the orthography *-s*, and the LRT for singular number would have one non-inflecting (orthographically null) LRI.

Each PC must specify what can be the input to it. This can be a general stem class such as any verb, a more specific stem class such as a transitive verb, or any defined PC or LRT. LRIs by default contribute an orthographic stem but no predicate value to the semantics. The adjectives library (Trimble, 2014) makes a slight change to this mechanism. As described in Trimble (2014), some nouns in some languages can be inflected with an adjective morphologically. When this is the case, the user is able to select a box on the morphotactics page indicating that adjective incorporation exists in the language. Upon

selecting this option, the user is then able to define incorporating-stem lexical rule types (IS-LRTs) within that PC. In these IS-LRTs, the user then defines one or more inflecting stem lexical rule instances for which they define the orthography as well as a predicate. The back end of the adjectives library then builds the EP for the adjective from this IS-LRT (along with its IS-LRT and PC) and identifies its ARG1 with the ARG0 of the EP for the noun it combines with. I draw heavily from the engineering done for incorporated adjectives, including the web-facing code that allows the user to define IS-LRTs and IS-LRTIs.

The case library (Drellishak, 2009) allows the user to define which values for the feature CASE are appropriate for the language in question. On the associated subpage for case, the user is able to choose from a list of case marking patterns including nominative/accusative, ergative/absolutive, tripartite, split-s, fluid-s, conditional split, and focus patterns. These patterns refer to how the core arguments (subject and object) of the verb are marked. Once defined, the user then interacts with other subpages to specify where these features appear in the language. Case marking can be done both affixally and adpositionally. If nouns are inflected for case, the user can define a PC which specifies case values on the noun itself. If instead case is marked with a preposition or postposition, the user can define a semantically empty adposition which specifies the case value on the noun.

The argument optionality library (Saleem, 2010; Saleem and Bender, 2010) allows the user to place specifications on the realization of arguments. If subjects or objects can be dropped in the language, the user can customize the rules which furthermore can account for this optionality by requiring, allowing, or forbidding the co-occurrence of agreement marking on the verb, or specifying contexts in which dropping is or is not allowed. When these choices are enabled, the customization system provides non-branching phrase structure rules in the grammar which satisfy the selectional requirements of the verb's valence lists without actually combining them with another element syntactically.

## **2.4 Summary**

In this chapter, I introduced the syntactic (HPSG) and semantic (MRS) frameworks I use in this thesis to give the context needed for understanding the specifics of the analyses I build. I also introduced the Grammar Matrix and its associated customization system. I provided an explanation of how this project is continuously refined with the addition of new libraries such as the one presented here and others which relate to noun incorporation. I demonstrated how libraries interact with one another, and how careful regression testing ensures compatibility.

## Chapter 3

### LITERATURE REVIEW

In this chapter, I present a typological survey of noun incorporation (NI) in the world’s languages and discuss existing syntactic and lexicalist analyses of the phenomenon. I also introduce literature on noun incorporation within the framework of HPSG and consider what parts of these analyses are applicable to my library. The chapter is organized as follows: In section 3.1, I provide definitions for terminology commonly used in this thesis. Section 3.2 presents a typological survey organized around answering the following questions about NI: *what can incorporate?* (3.2.1), *what does an incorporated noun look like?* (3.2.2), *what clausal effect does NI have?* (3.2.3), and *what should the semantic representation of NI look like?* (3.2.4). I then briefly discuss the schism in lexicalist v. syntactic approaches for analyzing NI in section 3.3. In section 3.4, I discuss the general classification of NI types in the literature and define my own grouping. Finally, section 3.5 describes three papers written on NI within the framework of HPSG to determine which parts of these analyses could be implemented within the Grammar Matrix customization system.

#### **3.1 Terminology**

I begin by defining some key terms in the paper. First, the phenomenon itself—noun incorporation. The definition of noun incorporation in the literature varies, with some encompassing more related constructions than others, but for the purposes of this paper and creating my library, I adopt Baker et al.’s (2005:135) definition which states that “**Noun Incorporation** (NI) is the phenomenon in which a nominal that would otherwise bear a grammatical relation to the verb (such as direct object) is expressed not as an independent noun phrase, but rather as a morphological root that is integrated into the inflected verb to

form a kind of composite form.” An **Incorporated Noun** (IN), then, is the nominal that is morphologically bound to the verb stem. I will also refer to **incorporating verbs** or NI constructions as verbs which have incorporated and IN.

Two related phenomena are **Pseudo Noun Incorporation** (PNI), also known as noun stripping (Gerdt, 2017), and **Obligatory Noun Incorporation**, also known as **Denominal Verbs** (DVs). PNI/noun stripping involves the syntactic juxtaposition of a noun and verb root, though external evidence (such as stress or other phonological indicators of wordhood) suggests that the noun does not form a word with the verb. Some authors will claim a language has NI, but represent incorporation as PNI (with the noun not morphologically attached to the verb) while other authors describing the same language do represent the INs as morphologically attached to the verb. This library supports INs as inflectional position classes on a verb root, and thus I will adopt the convention of treating nouns described as INs as affixes on the verb.

Denominal verbs are noun stems that are inflected with a verbal base, and the resulting complex word is a verb. These are similar to NI in that they involve the joining of a nominal element and verbal element and result in a complex verb, however, the root element for DVs is the noun, while the root element for NI is the verb. The verbalizing affixes in obligatory NI cannot stand alone (hence the terminology ‘verbal base’ rather than ‘verbal root’). Similarly to NI constructions, both the noun and verb components in DVs contribute semantically, and thus the verbal suffix is not simply a verbalizing suffix. However, due to the fact that DVs are in essence nominal inflection (rather than verbal inflection), I follow most analyses of NI and consider DVs to not be true forms of NI, and will not provide an analysis for them.

This thesis discusses the impact that NI has on the rest of the clause. Authors vary on the terminology used, ranging on whether NI may impact a verb’s argument structure, transitivity, or valence. I will use the first two interchangeably, but avoid using the third unless directly relevant to an author’s argument as valence has a different meaning than transitivity in HPSG. Additionally, I will use the terms **(in)transitive behavior** when talking about markers of transitivity such as (in)transitive morphemes, agreement morphology, or case

patterns. This notion of transitivity differs from the actual syntactic presence or absence of a direct object argument of the verb.

Finally, since NI deals with the morphological manifestation of nouns bearing a grammatical relation to the verb, this paper discusses verbal syntactic and semantic arguments. I refer to intransitive subjects as the S argument, and to transitive subjects and transitive direct objects as the A and O arguments respectively. I also refer to patient-like (as opposed to agent-like) S arguments, being those that are undergoers of the action rather than actors themselves.<sup>1</sup>

### 3.2 Typology of Noun Incorporation

In this section, I present the range of variation in noun incorporation in the languages of the world based on existing typological surveys of NI as well as typologically grounded analyses. This includes descriptions of what nouns can incorporate in relation to the verb, in what form they incorporate, what effect incorporation has on the rest of the clause, and how incorporated sentences vary in meaning from their unincorporated counterparts.

#### 3.2.1 What can incorporate?

Most commonly, the IN corresponds to what would have been the direct object of a transitive verb in a construction without noun incorporation. In the example from Mapudungun [arn] below, the direct object of the verb in (15a) *waka* ‘cow’ appears affixed to the verb in (15b).

- (15) a. *Ñi chao kintuley ta.chi pu waka*  
           *Ñi chao kintu-le-y ta.chi pu waka*  
           my father seek-PROG-IND.3SGS the COLL cow  
           ‘My father is looking for the cows.’ [arn] (Salas, 1992:195 via Baker et al., 2005:139)

---

<sup>1</sup>Throughout this paper, I standardize the glosses in IGT data to reference S, A, and O arguments for consistency.

- b. *Ñi chao kintuwakaley*  
 Ñi chao kintu-**waka**-le-y  
 My father seek-**cow**-PROG-IND.3SGS  
 ‘My father is looking for the cows.’ [arn] (Salas, 1992:195 via Baker et al., 2005:139)

Sometimes, patient-like intransitive subjects (16) and instrumental (17) or locative (18) oblique arguments can be incorporated. Each of these fall beyond the scope of my library.

- (16) *wa'tkaksahri'ne'*  
 wa'-t-ka-**ks**-a-hri'-ne'  
 FACT-DUP-NSGS-**dish**-Ø-shatter-PUNC  
 ‘The dish broke.’ [moh] (Baker, 1996:213 via Baker et al., 2005:140)
- (17) *ya' kikočillotete'ki panci*  
 ya' ki-**kočillo**-tete'ki panci  
 he (he)it-**knife**-cut bread  
 ‘He cut the bread with it (the knife).’ [nhe] (Merlan, 1976 via Mithun, 1984:861)
- (18) *gas.himtilílinau*  
 g-**as.him**-tilí-lin-au  
 PAST-**seashore**-walk.around-3-PL  
 ‘They walked around on the seashore.’ [kpy] (Bogoras, 1917 via Mithun, 1984:861)

Elements barred from incorporating include agent-like intransitive subjects as well as goal/malefactive/benefactive oblique arguments (Baker et al., 2005). (19a) shows that the object *athvno* ‘ball’ can incorporate, but (19b) shows that *wir* ‘baby’ cannot, unless it is understood to be the object rather than the goal. Similarly in (20), the only possible interpretation of the sentence with *waka* ‘cow’ incorporated is with it being the item that is bought (the object).

- (19) a. *ta'kheyathvnotsheru'* *ne owira'a*  
 t-a'-khey-**athvno**-tsher-u' ne owira'a  
 CIS-FACT-1SGS/FSGO-**ball**-NOML-give-PUNC NE baby  
 ‘I gave the ball to the baby.’ [moh] (Baker, 1996:207 via Baker et al., 2005:140)
- b. *\*ta'kewiru'* *ne athvno*  
 t-a'-ke-**wir**-u' ne athvno  
 CIS-FACT-1SGS/FSGO-**baby**-NOML-give-PUNC NE ball

NOT: ‘I gave the ball to the baby.’ (OK as ‘I gave the baby to the ball.’) [moh]  
 (Baker, 1996:207 via Baker et al., 2005:140)

- (20) *Juan ngillawakalelfiy (Pedro)*  
 Juan ngilla-**waka**-lel-fi-y (Pedro).  
 Juan buy-**cow**-BEN-3O-IND.3SGS Pedro  
 ‘Juan bought a cow for him (Pedro).’ (NOT: ‘Juan bought it for the cow.’) [arn]  
 (Baker et al., 2005:140)

### 3.2.2 What does an incorporated noun look like?

The most common form of noun incorporation involves a bare nominal. If an unincorporated noun appears with a determiner, numeral, case marking adposition, or is inflected for number, it will not when incorporated. In (21a) and (21b), the unincorporated noun *seuan* ‘man’ is marked with either a singular or plural number affix. When *seuan* is incorporated in (21c) and (21d), there is no nominal number marking affix present on the IN, and the its number value is interpreted by the verbal agreement affix. In (21a) and (21c), the agreement affix *ti-* indicates that the object is singular, and likewise the agreement affix *bi-* agrees with the plural number of the object in (21b) and (21d).<sup>2</sup>

- (21) a. *seuanide timuban*  
 seuan-ide ti-mu-ban  
 man-SG 1SGA.3SGO-see-PAST  
 ‘I saw the man’ [tix] (Allen et al., 1984 via Baker et al., 2005:141)
- b. *seuanin bimuban*  
 seuan-in bi-mu-ban  
 man-PL 1SGA.3PLO-see-PAST  
 ‘I saw the men’ [tix] (Allen et al., 1984 via Baker et al., 2005:141)
- c. *tiseuanmuban*  
 ti-seuan-mu-ban  
 1SGA.3SGO-see-PAST

---

<sup>2</sup>*ti-* and *bi-* are actually glossed as 1S/AO and 1S/BO in Baker et al.(2005). I’ve adapted the glossing to reflect the difference between intransitive and transitive subjects (S and A) and to highlight the number marked on the affix. A/B in Baker et al.’s gloss refer to different noun classes that in this case differ in number. See Baker et al., 2005:139 for more information.

‘I saw the man’ [tix] (Allen et al., 1984 via Baker et al., 2005:141)

d. *biseuanmuban*

bi-seuan-mu-ban

1SGA.3PLO-see-PAST

‘I saw the men’ [tix] (Allen et al., 1984 via Baker et al., 2005:141)

In some languages, the incorporated form of a noun has a different spelling than the unincorporated form (Aikhenvald, 2007). The Traditional Tiwi [tiw] sentence in (22) below incorporates *kiji* ‘stick’ and *maripi* ‘chest,’ but the unincorporated forms of these two nouns are *taɲini* and *pipwa* respectively (Aikhenvald, 2007:13 citing Osborne, 1974:49). This also shows an example of multiple nouns being incorporated, which is rare.

(22) *ngampirikijimaripirituwa*

nga-mpi-ri-kiji-maripi-rrituwa

we(INCL)-NP:her(DO)-CV-stick-chest-slit.open

‘We slit the chest [of a goose: FEM] with a stick.’ [tiw] (Lee, 1987:164 via Aikhenvald, 2007:13)

While the most common form of NI involves the incorporation of just a noun, in some languages, more noun phrase material can also be incorporated. For instance, Boumaa Fijian [fij] allows the incorporation of coordinated noun phrases (23) and modifiers (24) with INs.

(23) *e la’i taaniusebu’a*

e la’i taa-[niu-se-bu’a]

ASP go chop-[copra-or-firewood]

‘He’s gone to chop copra or firewood.’ [fij] (Dixon, 1988:227 via Aikhenvald, 2007:14)

(24) *‘anawacipo’i*

‘ana-[waci-po’i]

eat-[cooked.taro.leaves-rolled]

‘eat rolled taro leaves’ [fij] (Dixon, 1988:227 via Aikhenvald, 2007:14)

For this thesis, I consider the incorporation of more than a bare noun<sup>3</sup> and verbs with

---

<sup>3</sup>I do, however, account for the fact that the orthographic form of a noun when incorporated can differ from its free-standing form.

more than one IN outside the scope of the project.

### 3.2.3 *What clausal effect does noun incorporation have?*

In her influential 1984 paper, Mithun describes four types of noun incorporation. These four types exist in a hierarchy<sup>4</sup> such that if a language exhibits type III, it will also exhibit types I and II, and so on. She states that all languages which incorporate always have an equivalent, unincorporated paraphrase, and thus the choice to incorporate is driven by one of four purposes: lexicalizing common activities, foregrounding non-core arguments, backgrounding known discourse information, and semantically constraining arguments. These four types are outlined below.

#### **Type I: Lexical Compounding**

In this type, a noun root and a verb root, where the noun bears some semantic relationship to the verb, combine in the lexicon because the activity is common enough to be name-worthy. The incorporated noun here does not occur with any determiners, demonstratives, or sport any case marking that would appear in an unincorporated structure.

- (25) a. *Na'e haka 'e he siana 'a he ika*  
 Na'e haka 'e he siana 'a he ika  
 PAST cook ERG the man ABS the fish  
 'The man cooked a fish' [ton] (Chung, 1978:152 via Rosen, 1989:311)
- b. *Na'e hakaika 'a he siana*  
 Na'e haka-ika 'a he siana  
 PAST cook-fish ABS the man  
 'The man cooked fish' [ton] (Chung, 1978:152 via Rosen, 1989:311)

In the Tongan [ton] example in (25a) above, the unincorporated *ika* 'fish' both appears with a definite determiner and is marked for absolutive case. When it is incorporated in (25b), a bare form of the noun *ika* 'fish' appears affixed to the verb. Also of note is the fact that the subject *siana* 'man' is marked with ergative case in the unincorporated sentence,

---

<sup>4</sup>This is a different notion of hierarchy than the hierarchy of *types* described in the previous chapter.

but with absolutive case in the incorporated construction. This suggests that the incorporated construction is an intransitive clause where the verb has only one syntactic argument (the subject). However, not all lexical compounding results in a sentence with intransitive behavior. Compare (21) above.

### Type II: The Manipulation of Case

The function of type II is to incorporate a noun so that another argument can be promoted to the vacated position. Mithun states that oblique arguments can be promoted to direct object position for transitive O incorporation, or subject position for intransitive S incorporation. In this type, the transitivity of the clause is unaffected by incorporation. What can be promoted depends on the language, but is limited to oblique arguments such as locatives (26) and possessors (27). Notice the lack of prepositional/locative markers (*ičil*) for the promoted Yukatec Maya [yua] object *in-kool* ‘my cornfield’ in (26b).

- (26) a. *kinče'akk*                      *čee' ičeil inkool*  
 k-in-če'ak-Ø-k                      če' ičil in-kool  
 INCOMP-I-chop-it-IMPF tree **in my-cornfield**  
 ‘I chop the tree in my cornfield.’ [yua] (Bricker, 1978 via Mithun, 1984:858)
- b. *kinče'akčee'tik*                      *inkool*  
 k-in-če'ak-čee'-t-ik                      in-kool  
 INCOMP-I-chop-tree-TR-IMPF **my-cornfield**  
 ‘I clear my cornfield.’ [yua] (Bricker, 1978 via Mithun, 1984:858)
- (27) a. *tʉg ĩĩ*                      *datés*  
 tʉg ĩĩ                      da-tés  
 tooth **1SG-POSS** THEME-hurt  
 ‘My tooth hurts.’ [mbj] (Weir, 1990:323)
- b. *ĩĩh tʉgdatés*  
*ĩĩh tʉg-da-tés*  
**1SG** tooth-THEME-hurt  
 ‘I have toothache.’ (Lit. ‘I tooth-hurt’.) [mbj] (Weir, 1990:323)

In the Nadëb [mbj] example in (27b), the S argument *tʉg* ‘tooth’ is incorporated into the verb, and the possessor *ĩĩ* ‘my’ is promoted to subject position where it appears not as

a possessive pronoun, but rather as the personal pronoun  $\tilde{ih}$  ‘I’. Both the unincorporated and incorporated paraphrases are intransitive constructions. This type of NI is also known as **possessor raising**.

The same process happens for transitive clauses in Nadëb as in (28a) and (28b) where both the incorporated and unincorporated sentences are transitive.

- (28) a. **a**            *moo*h  $\tilde{ih}$  *hi*x $\tilde{i}$ t  
**a**            *moo*h  $\tilde{ih}$  *hi*-jx $\tilde{i}$ t  
**2SG+POSS** hand 1SG THEME+ASP-wash  
 ‘I wash your hands.’ [mbj] (Weir, 1990:324)
- b.  $\tilde{om}$   $\tilde{ih}$  *moo*h*hi*x $\tilde{i}$ t  
 $\tilde{om}$   $\tilde{ih}$  *moo*h*hi*-jx $\tilde{i}$ t  
**2SG** 1SG hand-THEME+ASP-wash  
 ‘I wash your hands.’ (Lit. ‘I hand-wash you.’) [mbj] (Weir, 1990:324)

Here, the noun *moo*h ‘hand’ is incorporated into the verb, and the possessor is promoted to the direct object. In (28b), the form of the second person pronoun changes from a possessive pronoun to a personal pronoun, but the person and number values remain the same, and the incorporated *moo*h ‘hand’ does belong to the second person referent.

### Type III: The Manipulation of Discourse Structure

For type III, the purpose of incorporation is to background information that was previously mentioned in discourse. In this type, NI can result in a decrease in transitivity as in the following example (like type I), or another argument can be promoted to O (or S) position (like type II). In other words, the structure of type III is not different than types I or II, but the motivation is different in that in the IN can only be a noun which is known from a previous sentence in the discourse, and its incorporation reduces its salience.

In (29b) from Huastla Nahuatl [nhe], *naka* ‘meat’ is incorporated (it was introduced previously in (29a)), and the resulting clause is intransitive.

- (29) a. *askeman ti*’kwa    ***nakatl***  
 askeman ti-’-kwa    ***nakatl***  
 never    you-it-eat **meat**  
 ‘You never eat meat.’ [nhe] (Merlan, 1976 via Mithun, 1984:860)

- b. *na' ipanima ninakakwa*  
 na' ipanima ni-**naka**-kwa  
 I always I-**meat**-eat  
 'I eat it (meat) all the time.' [nhe] (Merlan, 1976 via Mithun, 1984:861)

#### Type IV: Classificatory Noun Incorporation

This type involves the incorporation of a general noun which is accompanied by an external NP, and the purpose of incorporation is to narrow the scope of the NP. The IN and external NP are co-referential in that they represent the same entity. There are a few manifestations of this type of NI. Either a general noun can incorporate while a more specific noun remains outside of the verb, the same noun can appear incorporated and doubled, or some modifier of the IN appears stranded outside of the verb.

In the following example from Kunwinjku [gup], the IN *dulg* 'tree' is qualified by the unincorporated noun *mangaralaljmajn* 'cashew nut', which specifies what kind of tree was seen. This version of type IV noun incorporation is known elsewhere in the literature as **doubling**.

- (30) ... *benedulgnaŋ mangaralaljmajn*  
 ... bene-**dulg**-naŋ **mangaralaljmajn**  
 they.two-**tree**-saw **cashew.nut**  
 '... They saw a cashew tree.' [gup] (Oates, 1964 via Mithun, 1984:867)

In some cases, the doubled external NP does not need to be more specific than the IN as in the Rembarrnga [rmb] example in (31). It can never be, however, less specific than the IN (Rosen, 1989:297).

- (31) *kaʔaʔ parkaʔaʔtaŋjɪŋ*  
 kaʔaʔ-∅ par-kaʔaʔ-ta-ŋjɪŋ  
**paperbark**-NOM 3SG.O.3PL.TR.S-**paperbark**-stand-(CAUS)-PAST.CONT  
 'They would spread paperbark on the ground.' [rmb] (McKay, 1975:296 via Rosen, 1989:303)

Mithun claims that the external NP might not contain a noun but rather that other NP material can appear on their own. Elsewhere in the literature, this type IV noun incorpo-

ration is known as **stranding**. Stranded elements can include demonstratives as in (32) or adjectives as in (33), both from Caddo [cad].

- (32) *ná: kannúh 'a'*  
*ná: kan-núh-'a'*  
 that water-run.out-will  
 'That water will run out.' [cad] (Parks, 1977 via Mithun, 1984:865)

- (33) *wayah hákk 'uht 'í'sa'*  
*wayah hák-k'uht-'í'-sa'*  
 a.lot PROG-grass-be/grow-PROG  
 'There is a lot of grass.' [cad] (Chafe, 1976 via Mithun, 1984:866)

### 3.2.4 *What does the semantic representation of noun incorporation look like?*

Mithun's first three types provide some answer to the motivation behind noun incorporation. Lexical compounding is used to derive a word for a common or generalized activity. Case manipulation backgrounds the patient object of the verb to foreground an oblique argument, and discourse manipulation backgrounds ideas which have become old information in the context of a given discourse (not necessarily common activities). Often, but not always, the IN is non-referential or non-specific. Each of these reasons seems to take focus away from the incorporated noun, potentially to shift focus to a different argument. Indeed, incorporated nouns are said to be devoid of discourse focus (Gerdt, 2017). The fourth type of NI doesn't seem to impact the focus of the IN. Mithun states that the IN is used to narrow the scope of doubled material, though I believe an argument could be made that doubled and stranded material rather modifies the IN. Since Mithun states that all instances of NI have an equivalent unincorporated paraphrase, my main aim in modeling the semantic representation of NI will be to approximate the semantics of the corresponding sentences without incorporation. I do not provide a full information structure analysis of NI, leaving this to future work.

### ***3.3 Syntactic and Lexicalist Analyses of Noun Incorporation***

The issue of noun incorporation has sparked some debate in the literature over whether NI happens in the syntax or in the lexicon. The most well-known syntactic approach is a head movement analysis from Baker (1988) (See also an autolexical account of NI in Sadock, 1985). In the head movement analysis, the head noun of the NP which is the complement of a transitive verb moves to join directly with the verb. Stranding is due to the NP having additional material than the head noun, and the restriction that only the head can incorporate (though see examples (23) and (24) above).

Lexicalists argue that NI is a word-formation process that does not involve syntactic movement (Mithun, 1984; Di Sciullo and Williams, 1987; Rosen, 1989). Rosen (1989) argues that a morphological process combines the noun root and verb root, either saturating or qualifying the argument-taking property of the verb. This paper not only serves to argue for a lexicalist approach, but also introduces the notion that there are two distinct types of NI, and that a given language will exhibit one or the other. These two types are known as Compound/Compounding NI and Classifier NI. Most current work on NI from both lexicalist and syntactic schools of thought maintains this binary distinction (see for example Baker, 1996; Baker et al., 2005; Runner and Aranovich, 2003).

### ***3.4 Classifying Types of Noun Incorporation***

In this section, I discuss the the two types of NI introduced in the previous section, and use these observations as well as Mithun's four types to offer up my own classification of NI types which will support my analysis. The distinction between Compounding and Classifier NI has to do with evidence of the IN saturating the argument structure of the verb or not (Rosen, 1989). In Compounding NI, the process of incorporating the noun is seen to reduce the transitivity of the verb (i.e. turn a transitive verb into an intransitive one) because the IN satisfies one of the arguments of the verb. Classifier NI on the other hand does not impact the verb's transitivity. If a transitive verb seeks to take an object, the IN will not satisfy

this argument, and the verb will combine with another object.

Evidence for differences in transitivity marking can be seen in the following two examples from Southern Tiwa and Mapudungun.

- (34) a. *Tiseuanmuban*  
 Ti-seuan-mu-ban  
 1SGS/3SGO-man-see-PAST  
 ‘I saw the/a man.’ [tix] (Allen et al., 1984 via Baker et al., 2005:141)
- b. *Biseuanmuban*  
 Bi-seuan-mu-ban  
 1SGS/3PLO-man-see-PAST  
 ‘I saw the men.’ [tix] (Allen et al., 1984 via Baker et al., 2005:141)
- (35) a. *Ngillawaka(\*fi)n*  
 Ngilla-waka-(\*fi)-n  
 buy-cow-(\*3O)-IND.1SGS  
 ‘I bought a cow.’ [arn] (Baker et al., 2005:141)
- b. *ɲgillafiñ ti waka*  
 ɲgilla-fi-ñ ti waka  
 buy-3O-IND.1S the cow  
 ‘I bought the cow.’ [arn] (Baker et al., 2005:141)

In the Southern Tiwa [tix] examples in (34), the verbal affix which agrees with the subject and object indicates that it agrees with the incorporated noun just as it would with an unincorporated object (see (21a) and (21b) above). This suggests that the IN does not reduce the transitivity of the verb. The Mapudungun examples on the other hand show that the object marker which is present and agrees with the object in an unincorporated construction (35b) cannot be present when noun incorporation takes place (35a), suggesting that this type of NI does reduce the verb’s transitivity. The evidence for whether or not NI changes the verb’s transitivity for these two languages came from verbal agreement affixes, but could also come from case marking on core arguments or transitivity marking on the verb.

Classifier NI languages like Southern Tiwa—or Mohawk [moh] as in example (36) be-

low — allow stranded or doubled material to modify the incorporated noun, while Compounding NI languages like Mapudungun don't (37).

- (36) *Wa'katherahninu'*                      *thikv*  
 Wa'-k-ather-a-hninu-'                      thikv  
 FACT-1SGS-basket-Ø-buy-PUNC that  
 'I bought that basket.' [moh] (Baker, 1996:308 via Baker et al., 2005:140)
- (37) \**Pedro ngillawakay*                      *tüfachi*  
 Pedro ngilla-waka-y                      tüfachi  
 Pedro buy-cow-IND.3SGS that  
 'Pedro bought that cow.' [arn] (Baker et al., 2005:140)

Though prevalent in the literature, the distinction between Compounding and Classifier NI is not so clear cut. First, authors vary in how certain languages are classified, and which feature (either (in)transitive behavior or the presence of doubling) is the dividing axis. Baker (1996) classifies Southern Tiwa as a Compounding NI language (type III in his terms). The division in this paper is the ability for NI constructions to have doubled nouns. Southern Tiwa allows stranding but not doubling, thus he does not consider it to be a Classifier (type IV) language. However, Baker et al. (2005) do group Southern Tiwa with Classifier NI languages, using transitive agreement in incorporation constructions as the division factor, and easing the definition of Classifier NI languages to ones which *sometimes* allow doubling.

None of these analysis account for Mithun's type II noun incorporation in which some oblique argument is promoted to object position except for Baker et al. (2005) who treat possessor raising as a type of stranding which Compounding NI languages are able to do. Furthermore, work on Polynesian languages shows that Compounding NI languages such as Niuean, Maori, and Tongan do allow stranded modifiers (Massam, 2001, Chung and Ladusaw, 2004, Ball, 2005). These discrepancies, as well as additional insights from Mithun 1984 lead me to propose a language-agnostic three-way classification of the structural types of NI: Reduction NI, Promotion NI, and Modifier NI.

Reduction NI is structurally similar to Mithun's type I in that the overt transitivity of the sentence is reduced, though morphology may be consistent with either transitive or

intransitive sentences as in the examples (34a)-(35b). Promotion NI is structurally similar to Mithun’s type II where another element is promoted to object position. This element may be the possessor of the IN, or an oblique argument. Finally Modifier NI is structurally similar to Mithun’s type IV. This type encompass both stranding and doubling. As previously mentioned, Mithun’s type III is structurally the same as types I and II, differing in the fact that the IN is known from discourse. Thus, what she would classify as Type III NI would fall either into my Reduction NI or Promotion NI category based on the presence or absence of a promoted noun.

### **3.5 HPSG and Noun Incorporation**

Little work has been done in the HPSG literature to describe an analysis of noun incorporation.<sup>5</sup> In this section, I discuss three papers in this intersection, the first of which describes the different ways lexical rules can account for compound and classifier NI, the second of which describes pseudo noun incorporation in West Greenlandic (Kalaallisut) [kal], and the last of which describes two approaches to for modeling NI in Tongan [ton].

#### *3.5.1 “Noun incorporation and rule interaction in the lexicon” (Runner and Aranovich, 2003)*

This paper by Runner and Aranovich (2003) posits that there are two different types of lexical rules in lexicalist theories of grammar such as HPSG, and that these two types correspond with the differences between Compounding and Classifier NI. Their distinction originates in a pre-lexicalist analysis of word formation by Wasow (1977, 1980). The thought is that lexical rules have access to and modify lexical information. Syntactic rules — not to be confused with the phrase-structure rules of HPSG — on the other hand, have access to and modify syntactic or phrasal information. In terms of HPSG features, the lexical information modified would be found in CONT (semantic features), and the syntactic feature modified would be

---

<sup>5</sup>I came to this conclusion by searching through the HPSG handbook (Müller et al., 2024) and HPSG conference proceedings (<https://proceedings.hpsg.xyz/>)

ARG-ST (argument structure). The authors claim that the lexicalization of syntactic rules in HPSG misses the differences that used to be made distinct by having both lexical and syntactic rules in a grammar. Further, they argue that this distinction can be realized by differentiating between lexical rules that relate lexemes to lexemes, and those that relate words to words, the former mirroring Wasow’s lexical rules, and the latter his syntactic rules.

This paper follows three main HPSG assumptions that are slightly different from the version of HPSG used in the Grammar Matrix. They describe a Type-Based Derivational Morphology (TBDM) analysis where the input to a lexical rule type is a feature MORPH (analogous to the Matrix’s DTR). The phonological information provided by the input and contributed by the rule both appear as elements in the PHON list (analogous to the Matrix’s STEM). These translations between their analysis and the current framework are trivial, but other assumptions are not as easily incorporated.

The second assumption the paper makes is that there exists a distinction between *lexeme* and *word*. The authors argue that the *lexeme* type does not have syntactic features (i.e. ARG-ST) and that the type *word* must maintain the same semantic values inherited from the lexeme that fed the word. When lexemes become words, the ARG-ST list is predicted from the lexeme’s CONT (Davis and Koenig (2000, 2001)). Their main argument is that Compounding NI happens via a lexeme-to-lexeme (L-to-L) rule, while Classifier NI happens via a word-to-word (W-to-W) rule.

This paper distinguishes between Compounding NI and Classifier NI primarily on the basis of case changes in compounding NI languages. The argument is that Compounding NI reduces the verb’s length of the verb’s ARG-ST, but that the length remains the same for Classifier NI constructions. The other observation is that Compounding NI never allows doubling while Classifier NI does.

The general type described for both forms of NI is a binary rule that combines the signs of two stems in the input rather than affixing the phonological representation of a lexical rule to the stem of the daughter. In this way, the input is two distinct stems from the lexicon.

$$verb \left[ \begin{array}{l} PHON \quad [1] + [2] \\ MORPH \quad \langle verb [PHON [2]], noun [PHON [1]] \rangle \end{array} \right]$$

Figure 3.1: Noun Incorporation type constraint from Runner and Aranovich, 2003:366

This rule can be seen in Figure 3.1.

The noun and verb input signs to this binary rule could either be lexemes or words to account for either type of NI (the inputs for Compounding NI will be lexemes and the inputs for Classifier NI will be words). The resulting PHON value is a concatenation of the PHON values for the two inputs. The authors then provide two additional type constraints which both inherit from this common supertype for the two types of NI.

For Classifier NI, this subtype specifies that both signs in the input are words. This type can be seen in Figure 3.2. As a word, the verb here has access to an ARG-ST in which the second element, relating to the verb’s object, is an NP which shares its HEAD value with the second input word—the incorporated noun. The reasoning behind only co-indexing the HEAD value rather than the entire NP is that the noun input is a word while the element on the ARG-ST is a phrase. The output of this type has an ARG-ST list as well, though the second element is a type *non-canonical* NP (i.e. one that is not an overt, free-standing NP but rather realized as an affix or gap (Müller et al., 2024)) with the same semantic role as the NP on the input verb’s ARG-ST. Making this NP *non-canonical* allows it to be absent from its overt location in the syntax and still satisfy the argument structure requirements of the verb. The notion of transitivity in this paper is based off of the ARG-ST of the verb, rather than its valence lists. Markers of transitive behavior, such as case value and agreement morphology, are assigned based on the ARG-ST list.

For Classifier NI languages that do not strand or double, this type would be sufficient.

$$\text{verb} - \text{word} \left[ \begin{array}{l} \text{ARG-ST} \langle \boxed{1}, \text{NP}_{\text{non-canon}} : \boxed{2} \rangle \oplus \boxed{3} \\ \text{MORPH} \langle \text{verb} - \text{word} \left[ \begin{array}{l} \text{ARG-ST} \langle \boxed{1} \text{NP}, \text{NP} \left[ \text{HEAD} \boxed{5} \right] : \boxed{2} \rangle \oplus \boxed{3} L \right], \\ \text{noun} - \text{word} \left[ \text{SYN|HEAD} \boxed{5} \right] \end{array} \right], \rangle \end{array} \right]$$

Figure 3.2: Classifier NI type constraint from Runner and Aranovich, 2003:366

The authors do not provide an analysis for stranding or doubling, which is something I will account for in my own analysis.

For Compounding NI, the authors define a separate type constraint which can be seen in Figure 3.3. This type constraint accounts for the change in transitivity by first identifying the index of the incorporated noun with the semantic argument of the verb in a type definition, and second defining a new principle that impacts the linking principle (which links the lexeme's semantic arguments with elements on a word's ARG-ST). In this new principle called the Argument Saturation Convention, the authors posit that once a semantic argument of a verb is identified with a nominal index, it cannot be linked to an element in the ARG-ST. Thus when the incorporated nominal's index is identified with the undergoer feature of the verb's semantics, it is inaccessible to the verb's ARG-ST later on, and when the compound verb lexeme becomes a word by a later rule, the resulting ARG-ST does not contain an element that can be realized as a valence feature representing the object of the verb and the verb is therefore intransitive.

A piece of evidence used by the authors to distinguish the two types of NI into different lexical rules is the interaction with another lexical rule: applicative formation (AF). Since applicative formation alters the argument structure of the verb, they argue, it is a word-to-word rule given its need to access the ARG-ST feature of the verb. With the constraint that



Figure 3.3: Compounding NI type constraint from Runner and Aranovich, 2003:368

lexeme-to-lexeme rules all occur before word-to-word rules, both types of NI should be able to feed applicative rules, but applicative rules should only be able to feed Classifier NI. They provide evidence of NI feeding AF in both a Compounding and Classifier NI language, and claim to show that AF may only feed Classifier NI but not Compounding NI.

It does appear that AF can feed Classifier NI and vice versa in the Ainu [ain] examples below. In (38) the noun (the underlying object) is incorporated and then an applicative morpheme occurs which promotes an oblique object to direct object position. In (39), the applicative morpheme applies first, promoting *sapa* ‘head’ to direct object, and this object is subsequently incorporated into the noun.

- (38) *tamkurpoki akotametaye*  
tam-kurpoki a-ko-tam-etaye  
sword-bottom 1SG.A-APPL-sword-draw.out  
‘I drew the sword out from the bottom of the (other) sword.’ [ain] (Runner and Aranovich, 2003:372)

- (39) *ratki apa asapaepuni*  
ratki apa a-sapa-e-puni  
hung door 1SG.A-head-APPL-lift  
‘I lifted the suspended door with my head.’ [ain] (Runner and Aranovich, 2003:372)

However, the evidence for the order of AF and Compounding NI is less convincing. The Chukchi [ckt] examples below are meant to show that in (40a) the object *təkečʔə* ‘bait’ is incorporated, then an applicative process promotes the oblique argument *utkučʔ* ‘trap’ to

direct object position, evidenced by the absolutive case marking.

- (40) a. *ətləge utkučʔən təkəčʔəpəlanen*  
 ətləg-e utkučʔ-ən təkəčʔə=pəla-nen  
 father-ERG trap-ABS bait=leave-3SGS/3SG  
 ‘The father left the bait at the trap.’ [ckt] (Runner and Aranovich, 2003:372)
- b. *\*ətləgen təkəčʔa utkučʔəpəlagʔe*  
 ətləg-en təkəčʔ-a utkučʔə=pəla-gʔe  
 father-ABS bait-INSTR trap=leave-3SGS  
 ‘The father left the bait at the trap.’ [ckt] (Runner and Aranovich, 2003:372)

While a null applicative marker is certainly possible, the promotion of *utkučʔ* ‘trap’ can also be explained by the noun incorporation process itself similar to how other languages incorporate one argument and subsequently promote another to the vacated position. Similarly, (40b) is meant to show that an oblique argument promoted to object position by an applicative cannot then be incorporated. Again, there is no morphological evidence of an applicative, the claim being that an antipassive (null here but overt elsewhere) demoted *təkəčʔə* ‘bait’ to an instrumental argument, and a null applicative promoted *utkučʔ* ‘trap’. Then, the incorporation of *utkučʔ* ‘trap’ made the sentence ungrammatical. Another explanation for the ungrammaticality of this sentence is to say that Chukchi cannot incorporate oblique arguments.

### 3.5.2 “West Greenlandic noun incorporation in a monohierarchical theory of grammar”

(Malouf, 1999)

This paper by Malouf, 1999 provides an HPSG analysis of denominal verbs (DVs) in West Greenlandic (Kalaallisut) [kal]. While denominal verbs are not considered in my analysis, it is still beneficial to consider the arguments made here especially for stranding since this was not considered in Runner and Aranovich, 2003 above.

As described, a denominal verb is a complex verb formed from a noun root and a verbal base which is affixed to the noun root morphologically. Both the nominal and verbal components of DVs contribute semantic meaning, in other words the verbal affix does not

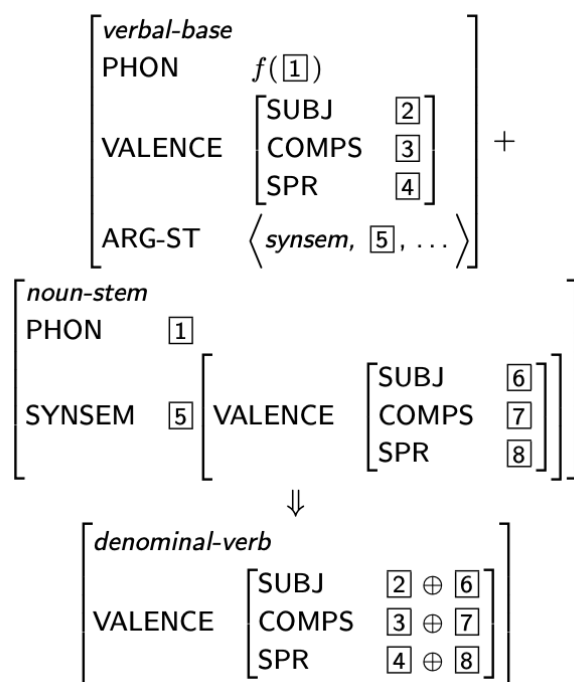


Figure 3.4: Noun Incorporation lexical rule from Malouf, 1999:9

just cause the verbalization of the noun. This is different than NI because instead of a root verb being inflected with (incorporating) a noun stem, DVs are noun roots inflected by a semantically contentful, verbalizing, bound suffix.

Malouf describes a binary lexical rule used to derive DVs in Kalaallisut which simply concatenates the valence lists of the two inputs. This rule is shown below in Figure 3.4. The phonological form of the DV is the PHON value of the noun after undergoing the function of the PHON value of the verbal base. Furthermore, the entire SYNSEM value of the noun is identified with the second argument on the verbal base’s ARG-ST list.

He provides the following example of the ‘incorporation’ of the noun stem *ammassak* ‘sardine’ into the verbal base *-tor* ‘eat.’ AVMS for these two lexical entries are shown in Figures 3.5 and 3.6. Each of these lexical entries inherit constraints from supertypes such as

<i>noun-stem</i>							
PHON	<i>&lt;ammassak&gt;</i>						
HEAD	<i>noun</i>						
VALENCE	<table style="border-collapse: collapse;"> <tr><td style="border-right: 1px solid black; padding-right: 5px;">SUBJ</td><td><i>&lt;&gt;</i></td></tr> <tr><td style="border-right: 1px solid black; padding-right: 5px;">COMPS</td><td><i>&lt;&gt;</i></td></tr> <tr><td style="border-right: 1px solid black; padding-right: 5px;">SPR</td><td><i>&lt;NP[erg]&gt;</i></td></tr> </table>	SUBJ	<i>&lt;&gt;</i>	COMPS	<i>&lt;&gt;</i>	SPR	<i>&lt;NP[erg]&gt;</i>
SUBJ	<i>&lt;&gt;</i>						
COMPS	<i>&lt;&gt;</i>						
SPR	<i>&lt;NP[erg]&gt;</i>						
CONTENT	<i>sardine_rel</i>						

Figure 3.5: A lexical entry for *ammassak*

‘sardine’

<i>verbal-base</i>							
PHON	<i>f<sub>tor</sub>(<span style="border: 1px solid black; padding: 0 2px;">1</span>)</i>						
HEAD	<i>verb</i>						
VALENCE	<table style="border-collapse: collapse;"> <tr><td style="border-right: 1px solid black; padding-right: 5px;">SUBJ</td><td><i>&lt;<span style="border: 1px solid black; padding: 0 2px;">2</span>&gt;</i></td></tr> <tr><td style="border-right: 1px solid black; padding-right: 5px;">COMPS</td><td><i>&lt;&gt;</i></td></tr> <tr><td style="border-right: 1px solid black; padding-right: 5px;">SPR</td><td><i>&lt;&gt;</i></td></tr> </table>	SUBJ	<i>&lt;<span style="border: 1px solid black; padding: 0 2px;">2</span>&gt;</i>	COMPS	<i>&lt;&gt;</i>	SPR	<i>&lt;&gt;</i>
SUBJ	<i>&lt;<span style="border: 1px solid black; padding: 0 2px;">2</span>&gt;</i>						
COMPS	<i>&lt;&gt;</i>						
SPR	<i>&lt;&gt;</i>						
ARG-ST	<i>&lt;<span style="border: 1px solid black; padding: 0 2px;">2</span>NP[abs]:<span style="border: 1px solid black; padding: 0 2px;">3</span>, <i>incorp</i> &amp; NP[inst]:<span style="border: 1px solid black; padding: 0 2px;">4</span>&gt;</i>						
CONTENT	<table style="border-collapse: collapse;"> <tr><td style="border-right: 1px solid black; padding-right: 5px;"><i>eat_rel</i></td><td></td></tr> <tr><td style="border-right: 1px solid black; padding-right: 5px;">ACTOR</td><td><span style="border: 1px solid black; padding: 0 2px;">3</span></td></tr> <tr><td style="border-right: 1px solid black; padding-right: 5px;">UNDERGOER</td><td><span style="border: 1px solid black; padding: 0 2px;">4</span></td></tr> </table>	<i>eat_rel</i>		ACTOR	<span style="border: 1px solid black; padding: 0 2px;">3</span>	UNDERGOER	<span style="border: 1px solid black; padding: 0 2px;">4</span>
<i>eat_rel</i>							
ACTOR	<span style="border: 1px solid black; padding: 0 2px;">3</span>						
UNDERGOER	<span style="border: 1px solid black; padding: 0 2px;">4</span>						

Figure 3.6: A lexical entry for *-tor* ‘-eat’

the ability for the noun stem to search for a specifier with ergative case and the fact that, as a bound verbal base, the *-tor*’s second element on the ARG-ST list is of type *incorp*, which is constrained to be *non-canonical*. *non-canonical* elements will not appear on the valence lists for verbs. This accounts for the empty COMPS list in the AVM.

From the lexical entry for *-tor*, we know that the ‘incorporated’ noun bears a semantic relationship with the verb, and appears as an element on its ARG-ST as a type *non-canonical* element. There is then a mismatch for verbal bases in the combined length of valence lists and ARG-ST. The IN is considered to be a part of the verb’s ARG-ST, but not a syntactic complement of the verb.

After these two signs go through the binary lexical rule in shown in Figure 3.4, the DV in Figure (3.7) is formed. The VALENCE lists for the DV are the concatenations of the VALENCE lists of the noun stem and the verbal base. Here, the one element on the DV’s SUBJ list comes from the verbal input, and the one element on the SPR list comes from the noun. The entire ARG-ST and CONTENT features are inherited from the verbal base, and likewise the HEAD value of the DV is *verb*.

To account for stranding, Malouf argues that the selectional criteria for modifiers by the ‘incorporated noun’ is inherited from the noun as it becomes a DV. The Adjunct lexical rule

PHON	$\langle ammassattor \rangle$						
HEAD	<i>verb</i>						
VALENCE	<table style="border-collapse: collapse;"> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;">SUBJ</td> <td style="padding: 2px 5px;"><math>\langle \boxed{2} \rangle</math></td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;">COMPS</td> <td style="padding: 2px 5px;"><math>\langle \rangle</math></td> </tr> <tr> <td style="border-right: 1px solid black; padding: 2px 5px;">SPR</td> <td style="padding: 2px 5px;"><math>\langle NP[erg] \rangle</math></td> </tr> </table>	SUBJ	$\langle \boxed{2} \rangle$	COMPS	$\langle \rangle$	SPR	$\langle NP[erg] \rangle$
SUBJ	$\langle \boxed{2} \rangle$						
COMPS	$\langle \rangle$						
SPR	$\langle NP[erg] \rangle$						
ARG-ST	$\langle \boxed{2} NP[abs]: \boxed{3}, NP[inst]: \boxed{4} \rangle$						
CONTENT	<table style="border-collapse: collapse;"> <tr> <td style="padding: 2px 5px;"><i>eat_rel</i></td> <td></td> </tr> <tr> <td style="padding: 2px 5px;">ACTOR</td> <td style="padding: 2px 5px;"><math>\boxed{3}</math></td> </tr> <tr> <td style="padding: 2px 5px;">UNDERGOER</td> <td style="padding: 2px 5px;"><math>\boxed{4} sardine\_rel</math></td> </tr> </table>	<i>eat_rel</i>		ACTOR	$\boxed{3}$	UNDERGOER	$\boxed{4} sardine\_rel$
<i>eat_rel</i>							
ACTOR	$\boxed{3}$						
UNDERGOER	$\boxed{4} sardine\_rel$						

Figure 3.7: The denominal verb *ammassattor* ‘sardine-eat’

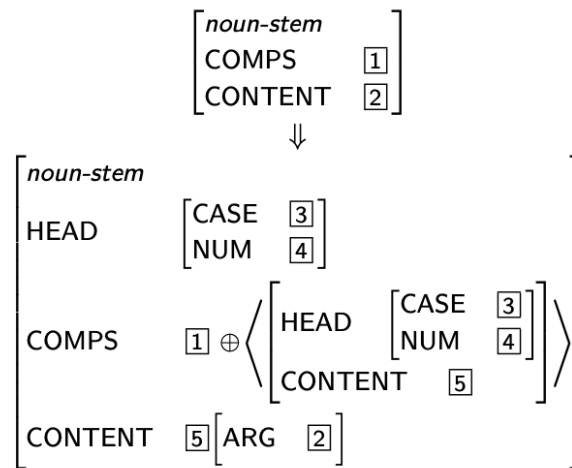


Figure 3.8: Adjunct lexical rule from (Malouf, 1999:11)

given by Malouf in Figure 3.8 shows how a noun stem looks for a modifier via its COMPS list, with constraints that the modifier have the same CASE and NUM values as the noun, and that the semantic content of the modifier should take scope over that of the noun.

A noun could first go through this adjunct lexical rule, then join with a verbal base to form a DV. This DV, by virtue of it inheriting the COMPS list of both its daughters, maintains the selectional constraints of nominal modifiers as the input noun. The entire DV then can take a valid modifier as a complement in a phrase structure rule such as shown in Figure 3.9 for the sentence in (41). Here, the element on the DV’s COMPS list (2) would have come from the comps list of the input noun.

- (41) *kissartumik kavvisurput*  
 kissartu-mik kavvi-sur-put  
 hot-INST coffee-drink-3PL.INDIC  
 ‘They drank hot coffee.’ [kal] (Fortescue, 1984:83 via Malouf, 1999:3)

To summarize, Malouf accounts for denominal verbs in Kalaallisut in a mixed-category manner by proposing a binary lexical rule which concatenates the valence lists of two inputs: a nominal root and a verbal base. Both inputs contribute to the semantics of the phrase, and the nominal displays a semantic relationship to the verb, but not a syntactic one. The composition of each input’s valence constraints accounts for the verbal syntactic properties of DVs (i.e. the ability to combine with absolutive subjects) as well as nominal properties (i.e. the ability to be modified or possessed).

### 3.5.3 “Tongan noun incorporation: Lexical sharing or argument inheritance” (Ball, 2005)

In this paper Ball (2005) argues for an Argument Inheritance Analysis for noun incorporation in Tongan [ton]. Incorporating verbs in Tongan pattern like intransitive verbs. When incorporation occurs, the subject is marked with absolutive case (43), but is marked with ergative case in standard transitive clauses (42).

- (42) *Na’e inu ‘a e kavá ‘e Sione*  
 Na’e inu ‘a e kavá ‘e Sione  
 PAST drink ABS DET kava.DEF ERG (name)

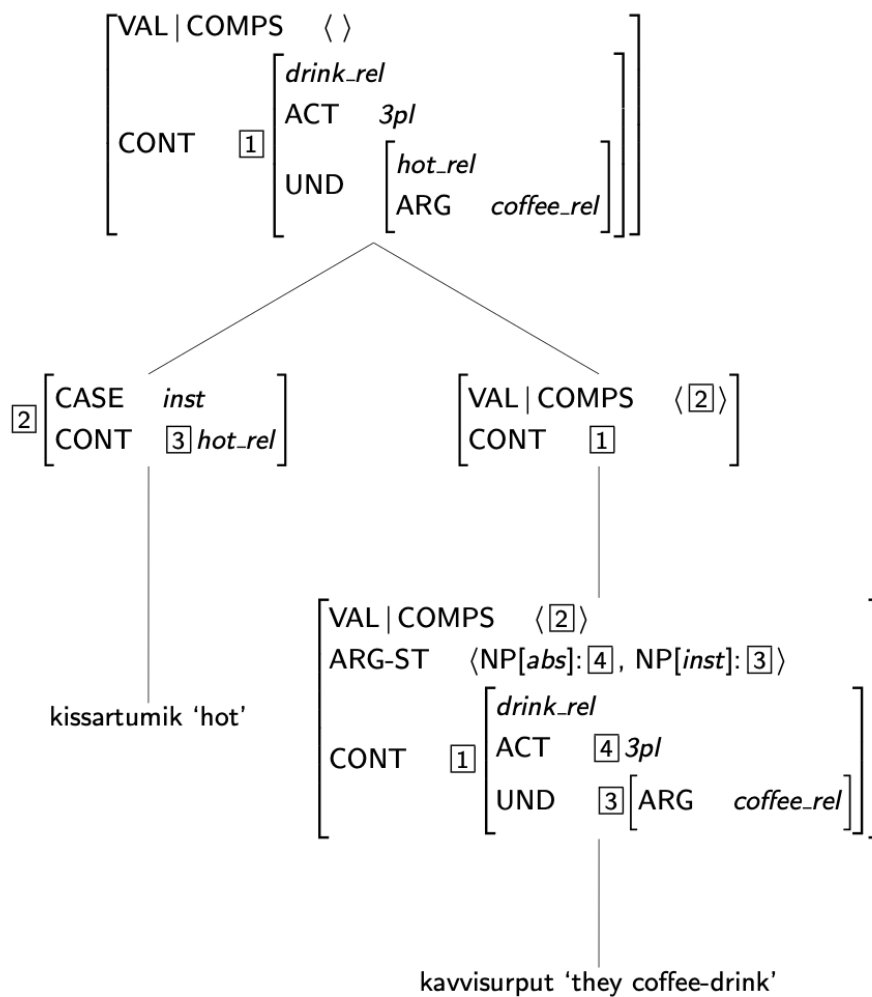


Figure 3.9: Structure of sentence (41)

‘Sione drank the kava.’ [ton] (Churchward, 1953:76 via Ball, 2005:9)

- (43) *Na‘e inu kava ‘a Sione*  
 Na‘e inu -kava ‘a Sione  
 PAST drink kava ABS (name)

‘Sione drank kava.’ [ton] (Churchward, 1953:76 via Ball, 2005:9)

Incorporated nouns in Tongan also appear to be accompanied by other NP elements such as the modifier *fo‘ou* ‘new’ in (44). In other words, one might think that full NPs are incorporated into the verb.

- (44) *Na‘e tā kītā fo‘ou ‘a Sione*  
 Na‘e tā -kītā fo‘ou ‘a Sione  
 PAST hit guitar new ABS (name)

‘Sione played a new guitar.’ [ton] (Churchward, 1953:76 via Ball, 2005:9)

However, Ball provides an alternative analysis in which the verb and the noun to form a word, and this complex word to form a phrase with any adnominals (also including prepositional phrases and relative clauses) that appear adjacent to the V+N complex. Thus, only the noun is incorporated, and other NP material appears outside of the verb. Evidence for treating the incorporating verb as a word but not including the adnominals in this word comes from a nominalizing affix that can nominalize a V+N complex (45a), but not a V+N+Adnominal complex (45b).

- (45) a. *inukava‘anga*  
 inu-kava-‘anga  
 drink-kava-NMLZ

‘place to drink kava’ [ton] (Churchward, 1953:76 via Ball, 2005:11)

- b. *\*fakataufalehinehina‘anga*  
 fakatau-fale-hinehina-‘anga  
 transact-house-white-NMLZ

Intended: ‘place for selling white houses’ [ton] (Churchward, 1953:76 via Ball, 2005:11)

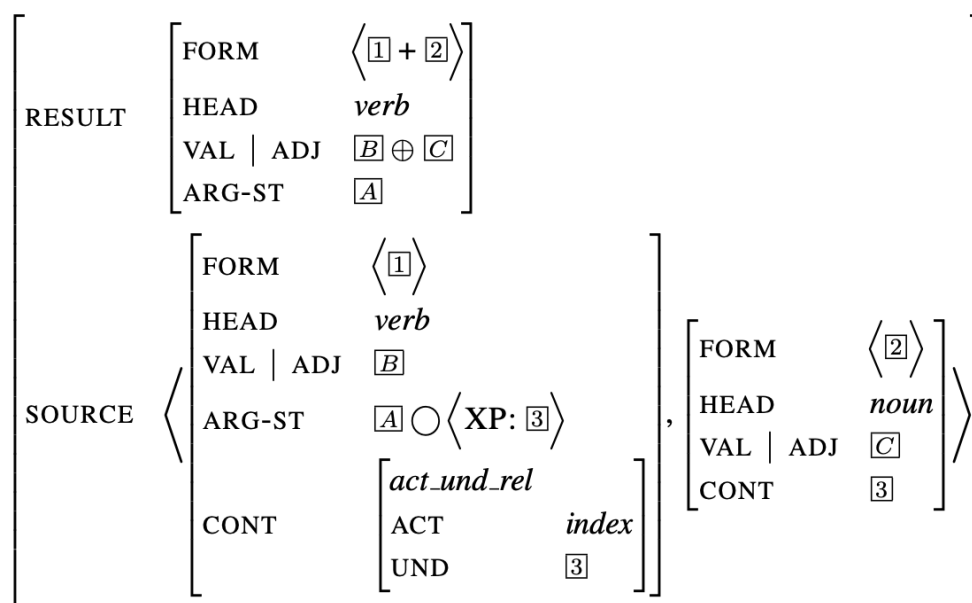


Figure 3.10: Noun Incorporation lexical rule from Ball, 2005:19

Similar to Malouf's (1999) binary-lexical rule analysis of Kalaallisut and Runner & Aranovich's (2003) binary TBDM rule, Ball proposes an Argument Inheritance Analysis of Tongan NI seen in Figure 3.10.

In this analysis, a binary lexical rule takes as input a noun stem and a verb stem, and the resulting HEAD value is *verb*. The rule concatenates the phonological forms of the two inputs, as well as the values on their ADJ lists. The noun root is co-indexed with a semantic argument of the verb (in CONT). The output of this lexical rule is a verb which has one fewer elements on its ARG-ST list than the input verb does. The excluded element is that which is co-indexed with the IN.

This rule is very similar to Runner & Aranovich's (2003) analysis of compound NI, which also identifies the IN with a semantic argument of the verb, and prohibits this element from being projected to the ARG-ST list of the output of the rule, resulting in a verb with reduced valence. The difference is that Runner and Aranovich constrain their rule to be an L-to-L

type rule, for which ARG-ST is not an applicable feature.<sup>6</sup> Ball does not place the same constraint on his rule.

#### 3.5.4 Summary

In summary, the three papers considered in this section analyze noun incorporation (Runner and Aranovich, 2003; Ball, 2005) and denominal verbs (Malouf, 1999) from the framework of HPSG. Each presents some type of binary lexical rule which combines properties of both the nominal and verbal elements of NI and DV constructions.

Runner and Aranovich account for the syntactic absence of the incorporated noun in one of two ways. In Compounding NI, the IN is barred from appearing on the ARG-ST of the complex verb. In Classifier NI, the IN is cast as type *non-cannonical*. Malouf (1999) and Ball (2005) take a more mixed-category approach to their analyses. Malouf concatenates the valence lists of the noun and verb elements, allowing the verb to select for its subject while also inheriting the nominal valence selectional criteria such as the ability to be modified. Ball's rule also inherits the valence information from the noun, but treats the rest of the output as decidedly verbal. The ARG-ST of the output is copied up from the input verb, except for the element co-indexed with the noun, along the lines of Runner & Aranovich's Compounding NI type.

None of these three proposed HPSG analyses of NI are implementable in the Grammar Matrix. All three involve not only the concatenation of the orthographic forms of the verb and IN, but also deal with the joining of two lexical entries into a single word which would serve as a terminal node in a tree structure. Based on how morphology is handled in the Matrix, I will need to introduce the orthography of the IN with an inflecting lexical rule instance rather than with a lexical entry.

Unlike Runner and Aranovich, 2003, the Grammar Matrix does not assume a distinction

---

<sup>6</sup>Recall that Runner and Aranovich argue that lexemes do not have the feature ARG-ST because the values of ARG-ST can be inferred from the lexeme's CONT feature when they become words via linking constraints.

between *lexeme* and *word* and cannot straight forwardly restrict some lexical rules to not have access to the ARG-ST feature. The Matrix also does not predict the ARG-ST of a verb based on its CONT values, instead this feature is always defined for the lexical entries of verbs. Additionally, the lexical entry for verbs already links the elements on the ARG-ST to elements on the valence lists. Thus, my analysis for NI won't entirely hinge on setting up the ARG-ST of the verb to not project an argument to the COMPS list, but rather will involve aiming for the desired behavior in conjunction with pre-defined valence lists.

While Runner and Aranovich 2003 do not provide an analysis of doubling or stranding, their insights into the differences between Compounding NI and Classifier NI are important to consider in my own analysis. They create two rules, one in which the IN is blocked from being realized on the verb's ARG-ST and thus blocked from being projected to the verb's COMPS list. The other rule realizes the IN on the verb's ARG-ST list as a *non-canonical* element which is blocked from being projected to the COMPS list. Both account for the absence of the IN in the syntax but in different ways. By leaving the *non-canonical* element on the ARG-ST, they are able to explain the transitive behavior of Classifier NI constructions, and likewise the intransitive behavior of Compounding NI due to the reduced length of the verb's ARG-ST. In the Matrix, there are other ways to configure the presence or absence of (in)transitive morphology and agreement affixes. Thus, focusing on the ARG-ST is not as important for my analysis. However, when morphemes or modifiers need to agree with the IN like in the Southern Tiwa examples in (21), it is useful to be able to reference it. In other words, I want to keep the IN on the verb's COMPS list but ensure that the verb does not overtly combine with a complement in the syntax, analogous to the *non-canonical* restriction laid out in their analysis.

I do not account for denominal verbs in my analysis, but Malouf's insights on stranded modifiers of DVs as well as Ball's analysis of stranding in Tongan are applicable. Both authors define an adjunct lexical rules which allow nouns to select for their modifiers. Since incorporating nouns and DVs are represented as mixed-category elements, they are also able to go through this lexical rule to combine with a modifier via the **head-comp** rule (Malouf,

1999) or the *head-adjuncts-schema* (Ball, 2005), and where the modifier acts semantically on the INDEX associated with the IN. This is possible in both architectures due to the complex verb inheriting the selectional constraints of the noun in a mixed-category analysis. I do not take this same approach, but do consider the need for stranded modifiers to combine with verbs in order to modify the correct element. My analysis needs to include a lexical rule for the incorporating verb which allows it to syntactically combine with a modifier of the IN.

Furthermore, while Runner and Aranovich (2003) and Ball (2005) provide HPSG analyses of NI within the binary division of Compounding and Classifier NI, both, in my opinion, lend credence to abandoning the strict classification in favor of different lexical rules for different manifestations of NI within a single language. Runner and Aranovich, 2003 use Ainu [ain] as an example of Classifier NI since it displays transitive behavior for NI constructions. However, as they point out, Ainu does not allow stranding or doubling which is typically a defining feature of that type of NI (Rosen, 1989, Baker et al., 2005). Ball (2005) on the other hand provides evidence of stranding in Tongan, which is considered to be a Compounding NI language (Rosen, 1989, Runner and Aranovich, 2003).

### **3.6 Summary**

In this chapter, I described the typology of noun incorporation cross-linguistically, looking specifically at what semantic roles can incorporate, the form of INs, and what additional elements can and cannot incorporate. I described what clausal impact NI has and how this is described in the literature as either a four-part or binary distinction. Finally, I considered what previous authors working within the framework of HPSG have written about NI, both broadly, and language-specifically. In the following chapter, I will provide my analysis of Reduction NI, Promotion NI, and Modifier NI.

## Chapter 4

### ANALYSIS

In this chapter, I discuss my syntactic and semantic analysis of Noun Incorporation in the frameworks of HPSG and MRS. For this analysis, I will be focusing on the incorporation of direct objects only, leaving subject and oblique argument incorporation to future work. Direct object incorporation represents by far the most frequent pattern, and providing an analysis for them is an important first step towards capturing the full picture of incorporation. While the analysis outlined in this chapter provides coverage for incorporating the element on the verb's COMPS list, future work could adapt the rules to instead target the element on the verb's SUBJ list, or other oblique arguments on the verb's COMPS list. I will also only be considering cases where the noun is the sole element incorporated (i.e. without determiners or productive morphology).

Based on my findings from my review of the literature, I described a three-way distinction between syntactic types of NI (see section 3.3). In the first, which I refer to as Reduction NI, a noun is simply incorporated into the verb which reduces the number of overt, unincorporated arguments of the verb. In the second type, which I refer to as Promotion NI, a noun is incorporated, and a different noun is promoted to the object position of the verb. Finally there are cases in which a noun is incorporated, and some element semantically related to the noun remains unincorporated. I refer to this type as Modifier NI. This element may be a co-referential noun that is more or as specific as the IN, or some modifying element such as a demonstrative or adjective.

This chapter is organized as follows: In section 4.1 I provide examples of Reduction NI, Promotion NI, and Modifier NI, comparing incorporating constructions to their unincorporated paraphrases when available in the languages to define the semantic representations I

wish to achieve for each type of NI. In section 4.2 I describe the lexical and phrase-structure rules needed to create the correct MRSs for these example sentences. I begin by presenting the first part of my analysis: a lexical rule shared by all types of NI (section 4.2.1). I then describe the second part of my analysis: separate lexical rules for Reduction NI (section 4.2.2) and Promotion NI (section 4.2.3), and a combination of lexical rules and unary phrase structure rules for stranded modifiers and doubled nouns for Modifier NI (section 4.2.4).

#### 4.1 *Worked Examples of Noun Incorporation*

In this section, I analyze six examples of NI found in the literature and provide the corresponding MRSs. The first two are of the Reduction NI type—with the first behaving intransitively and the second behaving transitively. The second two are of the Promotion NI type, showing possessor raising, and the promotion of an oblique argument to object position. Finally, I show two examples of Modifier NI: stranding and doubling.

##### 4.1.1 *Resulting intransitive verbs*

Reduction NI in some languages, such as Chukchi [ckt], results in a sentence that appears intransitive.

- (46) a. *ətləge qoraŋə təmnen*  
 ətləg-e qoraŋə-∅ təm-nen  
 father-ERG reindeer-ABS kill-3SGA.3SGO  
 ‘The father killed the/a reindeer.’ [ckt] (Polinsky, 1990 via Gerdtts, 2017:88)
- b. *ətləgen qaanməgʔe*  
 ətləg-en qaa-nmə-gʔe  
 father-ABS reindeer-kill-3SGS  
 ‘The father killed the reindeer.’ [ckt] (Polinsky, 1990 via Gerdtts, 2017:88)

Without incorporation, the transitive sentence shows ergative case marking on the subject and absolutive case marking on the object, and the agreement suffix on the verb agrees in person and number with both of these arguments. With incorporation, the subject is

marked with absolutive case, and the agreement suffix agrees with the subject but not the incorporated noun.

Both sentences in (46) should have the following semantic representation.

$$(47) \left[ \begin{array}{l} mrs \\ LTOP \quad \boxed{h1} \ h \\ INDEX \quad \boxed{e2} \ e \\ \\ RELS \quad \left\langle \begin{array}{l} \left[ \begin{array}{l} exist\_q\_rel \\ LBL \quad \boxed{h3} \ h \\ ARG0 \quad \boxed{x5} \ x \\ RSTR \quad \boxed{h6} \ h \\ BODY \quad \boxed{h4} \ h \end{array} \right], \left[ \begin{array}{l} \_father\_n\_rel \\ LBL \quad \boxed{h7} \ h \\ ARG0 \quad \boxed{x5} \end{array} \right], \left[ \begin{array}{l} \_kill\_v\_rel \\ LBL \quad \boxed{h8} \ h \\ ARG0 \quad \boxed{e2} \\ ARG1 \quad \boxed{x5} \\ ARG2 \quad \boxed{x9} \ x \end{array} \right], \\ \\ \left[ \begin{array}{l} \_reindeer\_n\_rel \\ LBL \quad \boxed{h10} \ h \\ ARG0 \quad \boxed{x9} \end{array} \right], \left[ \begin{array}{l} exist\_q\_rel \\ LBL \quad \boxed{h11} \ h \\ ARG0 \quad \boxed{x9} \\ RSTR \quad \boxed{h13} \ h \\ BODY \quad \boxed{h12} \ h \end{array} \right] \end{array} \right\rangle \\ \\ HCONS \quad \left\langle \left[ \begin{array}{l} qeq \\ HARG \quad \boxed{h6} \\ LARG \quad \boxed{h7} \end{array} \right], \left[ \begin{array}{l} qeq \\ HARG \quad \boxed{h13} \\ LARG \quad \boxed{h10} \end{array} \right] \right\rangle \end{array} \right]$$

The MRS in (47) shows a RELS list with two noun relations and their corresponding quantifiers. The ARG0 of the `_reindeer_n_rel`, which represents the INDEX of *qoraŋə* ‘reindeer,’ is identified with the ARG2 of the verb’s relation `_kill_v_rel`. This identification shows that *qoraŋə* ‘reindeer,’ incorporated or free-standing, should be linked with the semantic argument of the verb associated with it’s direct object. Each of the noun relations are quantified by a `_q_rel`. For the sentence in (46a), both `exist_q_rels` are contributed to the semantics by the `bare-np` rule. An analysis of NI thus needs to contribute this relation for the IN.

#### 4.1.2 Resulting transitive verbs

In some languages, such as Southern Tiwa [tix], Reduction NI does not affect the transitive behavior of the clause.<sup>1</sup> Consider the two sentences in (48a) and (48b).

- (48) a. *Tiseuanmuban*  
 Ti-seuan-mu-ban  
 1SGS/3SGO-man-see-PAST  
 ‘I saw the/a man’ [tix] (Allen et al., 1984 via Baker et al., 2005:141)
- b. *Biseuanmuban*  
 Bi-seuan-mu-ban  
 1SGS/3PLO-man-see-PAST  
 ‘I saw the men’ [tix] (Allen et al., 1984 via Baker et al., 2005:141)

In both sentences, the argument agreement affix agrees with both the subject and the incorporated noun, rather than only the subject.

An unincorporated paraphrase for (48a) is shown below.

- (49) *Tiseuanmuban*  
 Seuan-ide      Ti-mu-ban  
 man-SG      1SS/AO-see-PAST  
 ‘I saw the/a man’ [tix] (Allen et al., 1984 via Baker et al., 2005:141)

The MRS for (49), which is also the desired MRS for (48a), can be seen in (50).<sup>2</sup>

---

<sup>1</sup>This behavior is likewise true in Southern Tiwa when an oblique argument is promoted to object position. An agreement affix in such a construction would have three-way agreement with the subject, incorporated noun, and promoted noun. Because ditransitive verbs are not yet implemented in the Customization System, I do not provide an analysis for Promotion NI that would allow for such agreement patterns. This is left to future work.

<sup>2</sup>A feature of having identical MRS representations is that when the grammar is used in generation, one and the same MRS can give rise to both paraphrases, i.e. with and without NI.

$$(50) \left[ \begin{array}{l} mrs \\ LTOP \quad \boxed{h1} \ h \\ INDEX \quad \boxed{e2} \ e \\ \\ RELS \quad \left\langle \begin{array}{l} \left[ \begin{array}{l} exist\_q\_rel \\ LBL \quad \boxed{h3} \ h \\ ARG0 \quad \boxed{x5} \ x \\ RSTR \quad \boxed{h6} \ h \\ BODY \quad \boxed{h4} \ h \end{array} \right] , \left[ \begin{array}{l} \_man\_n\_rel \\ LBL \quad \boxed{h7} \ h \\ ARG0 \quad \boxed{x5} \end{array} \right] , \left[ \begin{array}{l} \_see\_v\_rel \\ LBL \quad \boxed{h8} \ h \\ ARG0 \quad \boxed{e2} \\ ARG1 \quad \boxed{x9} \ x \\ ARG2 \quad \boxed{x5} \end{array} \right] \right\rangle \\ \\ HCONS \quad \left\langle \begin{array}{l} \left[ \begin{array}{l} qeq \\ HARG \quad \boxed{h6} \\ LARG \quad \boxed{h7} \end{array} \right] \end{array} \right\rangle \\ \\ ICONS \quad \left\langle \begin{array}{l} \left[ \begin{array}{l} non-focus \\ LARG \quad \boxed{e2} \\ RARG \quad \boxed{x9} \end{array} \right] \end{array} \right\rangle \end{array} \right]$$

As in the intransitive Reduction NI example above, the ARG2 of the verb's relation is identified with the ARG0 of the noun relation for the object/IN.<sup>3</sup> Southern Tiwa has argument optionality, so the unexpressed subject in (49) appears in the MRS as a *non-focus* element on the ICONS list.

#### 4.1.3 Possessor raising

Possessor raising refers to the promotion of a noun's possessor to direct object position when that noun is incorporated. An example of possessor raising can be seen in the Mapudungun [arn] sentence in (51).

- (51) *Juan ngillawakafiy Pedro*  
 Juan ngilla-waka-fi-y Pedro  
 Juan buy-cow-3O-IND.3SGS Pedro  
 'Juan bought Pedro's cow.' [arn] (Baker et al., 2005:167)

<sup>3</sup>As will be explained in section (4.2.2), the MRS for the incorporation construction in (48a) actually differs slightly from that in (50). Its ICONS list has one more *non-focus* relation whose RARG is identified with the ARG0 of the IN.

Here, the promoted object *Pedro* has a possessive relationship with the incorporated noun *waka*, ‘cow’.

If the possessive noun phrase *Pedro ñi waka* ‘Pedro’s cow’ were the object of the verb, the MRS would look like the following:<sup>4</sup>

$$(52) \left[ \begin{array}{l} \text{mrs} \\ \text{LTOP} \quad \boxed{h1} \ h \\ \text{INDEX} \quad \boxed{e2} \ e \\ \\ \text{RELS} \quad \left\langle \begin{array}{l} \left[ \begin{array}{l} \text{exist\_q\_rel} \\ \text{LBL} \quad \boxed{h3} \ h \\ \text{ARG0} \quad \boxed{x5} \ x \\ \text{RSTR} \quad \boxed{h6} \ h \\ \text{BODY} \quad \boxed{h4} \ h \end{array} \right] , \left[ \begin{array}{l} \text{\_juan\_n\_rel} \\ \text{LBL} \quad \boxed{h7} \ h \\ \text{ARG0} \quad \boxed{x5} \end{array} \right] , \left[ \begin{array}{l} \text{\_buy\_v\_rel} \\ \text{LBL} \quad \boxed{h8} \ h \\ \text{ARG0} \quad \boxed{e2} \\ \text{ARG1} \quad \boxed{x5} \\ \text{ARG2} \quad \boxed{x9} \ x \end{array} \right] , \\ \\ \left[ \begin{array}{l} \text{exist\_q\_rel} \\ \text{LBL} \quad \boxed{h10} \ h \\ \text{ARG0} \quad \boxed{x9} \\ \text{RSTR} \quad \boxed{h12} \ h \\ \text{BODY} \quad \boxed{h11} \ h \end{array} \right] , \left[ \begin{array}{l} \text{poss\_rel} \\ \text{LBL} \quad \boxed{h13} \ h \\ \text{ARG0} \quad \boxed{e14} \ e \\ \text{ARG1} \quad \boxed{x9} \\ \text{ARG2} \quad \boxed{x15} \ x \end{array} \right] , \left[ \begin{array}{l} \text{exist\_q\_rel} \\ \text{LBL} \quad \boxed{h16} \ h \\ \text{ARG0} \quad \boxed{x15} \\ \text{RSTR} \quad \boxed{h18} \ h \\ \text{BODY} \quad \boxed{h17} \ h \end{array} \right] \right\rangle \\ \\ \left[ \begin{array}{l} \text{\_pedro\_n\_rel} \\ \text{LBL} \quad \boxed{h19} \ h \\ \text{ARG0} \quad \boxed{x15} \end{array} \right] , \left[ \begin{array}{l} \text{\_cow\_n\_rel} \\ \text{LBL} \quad \boxed{h13} \\ \text{ARG0} \quad \boxed{x9} \end{array} \right] \\ \\ \text{HCONS} \quad \left\langle \left[ \begin{array}{l} \text{qeq} \\ \text{HARG} \quad \boxed{h6} \\ \text{LARG} \quad \boxed{h7} \end{array} \right] , \left[ \begin{array}{l} \text{qeq} \\ \text{HARG} \quad \boxed{h12} \\ \text{LARG} \quad \boxed{h13} \end{array} \right] , \left[ \begin{array}{l} \text{qeq} \\ \text{HARG} \quad \boxed{h18} \\ \text{LARG} \quad \boxed{h19} \end{array} \right] \right\rangle \end{array} \right]$$

In this MRS, the ARG2 of the verb’s relation is identified with the ARG0 of *\\_cow\_n\_rel* which is the head noun, or possessum, of the possessive noun phrase. The *poss\_rel* EP relates the possessor and possessum by identifying their ARG0s with the relations ARG2 and ARG1 respectively. One goal for modeling possessor raising is to replicate the semantics of possession using the *poss\_rel* EP following that of Nielsen (2018) and the English Resource Grammar (Flickinger 2000, 2011).

<sup>4</sup>An unincorporated paraphrase for (51) was not given in Baker et al. (2005). This phrase is based on the description of nominal possession in Smeets (Smeets, 2008).

## 4.1.4 Oblique argument promotion

Promoted objects are not always the possessors of the IN, as evidenced by the Mohawk [moh] example in (53). Here, the promoted object is a benefactive argument.

- (53) *Ta'kheyathvnotsheru'* *ne owira'a.*  
 T-a'-khey-athvno-tsher-u-' ne owira'a.  
 CIS-FACT-1SGS/FSGO-ball-NMLZ-give-PUNC NE baby  
 'I gave the ball to the baby.' [moh] (Baker, 1996:207 via Baker et al., 2005:140)

An unincorporated paraphrase of (53) is not given in Baker et al. (2005), however the desired semantics are shown below.

- (54) 
$$\left[ \begin{array}{l} mrs \\ \text{LTOP} \quad \boxed{h1} \ h \\ \text{INDEX} \quad \boxed{e2} \ e \\ \\ \text{RELS} \quad \left\langle \begin{array}{l} \left[ \begin{array}{l} \_give\_v\_rel \\ \text{LBL} \quad \boxed{h3} \ h \\ \text{ARG0} \quad \boxed{e2} \\ \text{ARG1} \quad \boxed{x4} \ x \\ \text{ARG2} \quad \boxed{x5} \ x \end{array} \right], \left[ \begin{array}{l} \_ball\_n\_rel \\ \text{LBL} \quad \boxed{h6} \ h \\ \text{ARG0} \quad \boxed{x5} \end{array} \right], \left[ \begin{array}{l} exist\_q\_rel \\ \text{LBL} \quad \boxed{h7} \ h \\ \text{ARG0} \quad \boxed{x5} \\ \text{RSTR} \quad \boxed{h9} \ h \\ \text{BODY} \quad \boxed{h8} \ h \end{array} \right], \\ \\ \left[ \begin{array}{l} benefactive\_rel \\ \text{LBL} \quad \boxed{h3} \ h \\ \text{ARG0} \quad \boxed{e11} \ e \\ \text{ARG1} \quad \boxed{e2} \\ \text{ARG2} \quad \boxed{x12} \ x \end{array} \right], \left[ \begin{array}{l} exist\_q\_rel \\ \text{LBL} \quad \boxed{h13} \ h \\ \text{ARG0} \quad \boxed{x12} \\ \text{RSTR} \quad \boxed{h15} \ h \\ \text{BODY} \quad \boxed{h14} \ h \end{array} \right], \left[ \begin{array}{l} \_baby\_n\_rel \\ \text{LBL} \quad \boxed{h16} \ h \\ \text{ARG0} \quad \boxed{x12} \end{array} \right] \end{array} \right\rangle \\ \\ \text{HCONS} \quad \left\langle \left[ \begin{array}{l} qeq \\ \text{HARG} \quad \boxed{h9} \\ \text{LARG} \quad \boxed{h6} \end{array} \right], \left[ \begin{array}{l} qeq \\ \text{HARG} \quad \boxed{h15} \\ \text{LARG} \quad \boxed{h16} \end{array} \right] \right\rangle \\ \\ \text{ICONS} \quad \left\langle \left[ \begin{array}{l} non-focus \\ \text{LARG} \quad \boxed{e2} \\ \text{RARG} \quad \boxed{x4} \end{array} \right] \right\rangle \end{array} \right]$$

The ARG0 of `_ball_n_rel` contributed by the IN is identified with the ARG2 of the verb's relation. In order to account for the oblique argument, *owira'a* 'baby,' I follow Curtis (2018) in creating an EP for oblique arguments in which the verb's ARG0 is identified with the relation's ARG1, and the oblique argument's ARG0 is identified with the relation's ARG2.

#### 4.1.5 Doubling

Doubling occurs when a general noun is incorporated into the verb, and a co-referential, more specific (or as specific) noun remains outside of the verb and behaves as the direct object of the verb. An example from Mohawk [moh] is given below.

- (55) *Sha'teku nikuti rabahbót wahv[i]tsyahninu' ki rake'niha*  
 Sha'teku ni-kuti rabahbót wa-hv-[i]tsy-a-hninu-'  
 eight PART-ZPLS bullhead FACT-MSG-fish-Ø-buy-PUNC this my-father  
 'My father (fish-)bought eight bullheads.' [moh] (Baker, 1996:310 via Baker et al., 2005:149)

With doubling, there is no unincorporated paraphrase available. In my analysis, I strive to design a semantic representation for the sentence which indicates that the IN is a semantic argument of the verb, and that the doubled noun modifies the meaning of the IN in some way. I provide the following MRS for (55).

$$(56) \left[ \begin{array}{l} \text{mrs} \\ \text{LTOP } \boxed{h1} \ h \\ \text{INDEX } \boxed{e2} \ e \\ \\ \text{RELS } \left\langle \begin{array}{l} \left[ \begin{array}{l} \text{type\_rel} \\ \text{LBL } \boxed{h3} \ h \\ \text{ARG0 } \boxed{e4} \ e \\ \text{ARG1 } \boxed{x6} \ x \\ \text{ARG2 } \boxed{x5} \ x \end{array} \right], \left[ \begin{array}{l} \text{exist\_q\_rel} \\ \text{LBL } \boxed{h7} \ h \\ \text{ARG0 } \boxed{x5} \\ \text{RSTR } \boxed{h9} \ h \\ \text{BODY } \boxed{h8} \ h \end{array} \right], \left[ \begin{array}{l} \text{\_eight\_a\_rel} \\ \text{LBL } \boxed{h10} \ h \\ \text{ARG0 } \boxed{e11} \ e \\ \text{ARG1 } \boxed{x5} \end{array} \right], \\ \\ \left[ \begin{array}{l} \text{\_bullhead\_n\_rel} \\ \text{LBL } \boxed{h10} \\ \text{ARG0 } \boxed{x5} \end{array} \right], \left[ \begin{array}{l} \text{\_buy\_v\_rel} \\ \text{LBL } \boxed{h12} \ h \\ \text{ARG0 } \boxed{e2} \\ \text{ARG1 } \boxed{x13} \ x \\ \text{ARG2 } \boxed{x6} \end{array} \right], \left[ \begin{array}{l} \text{exist\_q\_rel} \\ \text{LBL } \boxed{h14} \ h \\ \text{ARG0 } \boxed{x6} \\ \text{RSTR } \boxed{h16} \ h \\ \text{BODY } \boxed{h15} \ h \end{array} \right], \\ \\ \left[ \begin{array}{l} \text{\_fish\_n\_rel} \\ \text{LBL } \boxed{h3} \\ \text{ARG0 } \boxed{x6} \end{array} \right], \left[ \begin{array}{l} \text{\_father\_n\_rel} \\ \text{LBL } \boxed{h17} \ h \\ \text{ARG0 } \boxed{x13} \end{array} \right], \left[ \begin{array}{l} \text{pron\_rel} \\ \text{LBL } \boxed{h18} \ h \\ \text{ARG0 } \boxed{x19} \ x \end{array} \right], \\ \\ \left[ \begin{array}{l} \text{exist\_q\_rel} \\ \text{LBL } \boxed{h20} \ h \\ \text{ARG0 } \boxed{x13} \\ \text{RSTR } \boxed{h22} \ h \\ \text{BODY } \boxed{h21} \ h \end{array} \right], \left[ \begin{array}{l} \text{poss\_rel} \\ \text{LBL } \boxed{h17} \\ \text{ARG0 } \boxed{e23} \ e \\ \text{ARG1 } \boxed{x13} \\ \text{ARG2 } \boxed{x19} \end{array} \right], \left[ \begin{array}{l} \text{exist\_q\_rel} \\ \text{LBL } \boxed{h24} \ h \\ \text{ARG0 } \boxed{x19} \\ \text{RSTR } \boxed{h26} \ h \\ \text{BODY } \boxed{h25} \ h \end{array} \right] \end{array} \right\rangle \\ \\ \text{HCONS } \left\langle \left[ \begin{array}{l} \text{qeq} \\ \text{HARG } \boxed{h9} \\ \text{LARG } \boxed{h10} \end{array} \right], \left[ \begin{array}{l} \text{qeq} \\ \text{HARG } \boxed{h16} \\ \text{LARG } \boxed{h3} \end{array} \right], \left[ \begin{array}{l} \text{qeq} \\ \text{HARG } \boxed{h22} \\ \text{LARG } \boxed{h17} \end{array} \right], \left[ \begin{array}{l} \text{qeq} \\ \text{HARG } \boxed{h26} \\ \text{LARG } \boxed{h18} \end{array} \right] \right\rangle \end{array} \right]$$

The EP `type_rel` is a relation in the semantics which relates the IN and the doubled noun. The ARG1 of the `type_rel` is identified with the ARG0 of the IN (`_fish_n_rel`), and the ARG2 of the `type_rel` is identified with the ARG0 of the doubled noun (`_bullhead_n_rel`). As with other modifiers of a noun (i.e. adjectives), the modifier of the IN should have the same scopal level as the IN. Thus, in the MRS, the noun relation for the IN shares its LBL with the `type_rel`.

#### 4.1.6 Stranding

Modifiers of the IN may appear outside of the verb as in the following example from Southern Tiwa [tix].

- (57) a. *Yede dirude ak'arhi*  
 Yed-e diru-ide a-kar-hi  
 that-SG chicken-SG 2SG.AO-eat-FUT  
 'You will eat that chicken.' [tix] (Allen et al., 1984:296)
- b. *yede adirukarhi*  
 yed-e a-diru-kar-hi  
 that-SG 2SGS.AO-chicken-eat-FUT  
 'You will eat that chicken.' [tix] (Allen et al., 1984:296 via Baker et al., 2005:150)

In the unincorporated paraphrase in (57a), *yede dirude* 'that chicken' forms a noun phrase. Cross-linguistically, demonstratives can be pronouns, adjectives, or adverbs (Levinson, 2018). Furthermore, in UW's LING 567 Grammar Engineering course, adnominal demonstratives can be modeled as either adjectives or determiners.<sup>5</sup> The type of demonstrative that I handle in this analysis is the adjective type, for which the predication contributes an `_a_rel` to the semantics. In HPSG, the modified noun appears on the adjective's MOD list, which unlike the valence lists (SUBJ, COMPS, etc.), is a value of the HEAD feature. The MRS for both (57a) and (57b) is given below.

---

<sup>5</sup><https://courses.washington.edu/ling567/testsuites.html#adj>

$$(58) \left[ \begin{array}{l} \text{mrs} \\ \text{LTOP} \quad \boxed{h1} \ h \\ \text{INDEX} \quad \boxed{e2} \ e \\ \\ \text{RELS} \quad \left\langle \begin{array}{l} \left[ \begin{array}{l} \text{exist\_q\_rel} \\ \text{LBL} \quad \boxed{h3} \ h \\ \text{ARG0} \quad \boxed{x5} \ x \\ \text{RSTR} \quad \boxed{h6} \ h \\ \text{BODY} \quad \boxed{h4} \ h \end{array} \right] , \left[ \begin{array}{l} \text{\_that\_a\_rel} \\ \text{LBL} \quad \boxed{h7} \ h \\ \text{ARG0} \quad \boxed{e8} \ e \\ \text{ARG1} \quad \boxed{x5} \end{array} \right] , \left[ \begin{array}{l} \text{\_eat\_v\_rel} \\ \text{LBL} \quad \boxed{h9} \ h \\ \text{ARG0} \quad \boxed{e2} \\ \text{ARG1} \quad \boxed{x10} \ x \\ \text{ARG2} \quad \boxed{x5} \end{array} \right] \\ \\ \left[ \begin{array}{l} \text{\_chicken\_n\_rel} \\ \text{LBL} \quad \boxed{h7} \\ \text{ARG0} \quad \boxed{x5} \end{array} \right] \end{array} \right\rangle \\ \\ \text{HCONS} \quad \left\langle \begin{array}{l} \left[ \begin{array}{l} \text{qeq} \\ \text{HARG} \quad \boxed{h6} \\ \text{LARG} \quad \boxed{h7} \end{array} \right] \end{array} \right\rangle \\ \\ \text{ICONS} \quad \left\langle \begin{array}{l} \left[ \begin{array}{l} \text{non-focus} \\ \text{LARG} \quad \boxed{e2} \\ \text{RARG} \quad \boxed{x10} \end{array} \right] \end{array} \right\rangle \end{array} \right]$$

In this MRS, the modifier relation (`\_that\_a\_rel`) shares its ARG1 with the ARG0 of the noun relation (`\_chicken\_n\_rel`) that it modifies. These two relations also identify their LBLs because their scopes are equal.

#### 4.1.7 Summary

In this section, I illustrated and described what the semantics of NI should look like across the variation of Reduction NI, Promotion NI, and Modifier NI. In general, the MRS of an incorporation construction should be identical to that of its unincorporated paraphrase, when available in the language. In the following section, I outline the rules needed to produce each of the MRSs above.

## 4.2 Lexical and Phrase Structure Rules of Noun Incorporation

In this section, I walk through the lexical and phrase structure rules which make up my analysis of the three types of NI. I begin by describing a lexical rule shared by all three types, then introduce secondary rules for Reduction NI, Promotion NI, and Modifier NI. Additionally, I present two unary phrase structure rules for Modifier NI which allow an external element to modify the IN.

### 4.2.1 A common noun incorporation lexical rule

Each of the varieties of NI introduced above similarly involves the morphological incorporation of the noun. I analyze this preliminary process identically across all three types. Since the noun is affixed to the verb, I define a verbal lexical rule that may apply to transitive verbs and orthographically affixes the noun. An example of such a rule can be seen in (59) below.

$$(59) \left[ \begin{array}{l} \textit{noun-incorp-lex-rule} \\ \\ \text{SYNSEM ... COMPS} \left[ \begin{array}{l} \text{FIRST} \left[ \begin{array}{l} \text{LOCAL.CONT.HOOK} \left[ \begin{array}{l} \text{INDEX} \quad \boxed{1} \\ \text{LTOP} \quad \quad \boxed{3} \end{array} \right] \\ \text{REST} \quad \boxed{4} \end{array} \right] \end{array} \right] \\ \\ \text{DTR.SYNSEM ... COMPS.REST} \quad \boxed{4} \\ \\ \text{C-CONT} \left[ \begin{array}{l} \text{HCONS} \left\langle \begin{array}{l} \textit{qeq} \\ \text{HARG} \quad \boxed{2} \\ \text{LARG} \quad \boxed{3} \end{array} \right\rangle \\ \\ \text{RELS} \left\langle \begin{array}{l} \left[ \begin{array}{l} \textit{noun-relation} \\ \text{PRED} \quad \text{"\_house\_n\_rel"} \\ \text{ARG0} \quad \quad \boxed{1} \\ \text{LBL} \quad \quad \quad \boxed{3} \end{array} \right] , \\ \left[ \begin{array}{l} \textit{quant-relation} \\ \text{PRED} \quad \text{"exist\_q\_rel"} \\ \text{ARG0} \quad \quad \boxed{1} \\ \text{RSTR} \quad \quad \quad \boxed{2} \end{array} \right] \end{array} \right\rangle \end{array} \right] \end{array} \right] \end{array} \right]$$

This rule contributes the necessary semantic information of the noun by defining a *noun-relation*, a *quant-relation*, and the necessary *qeq* for representing scope. In sentences without NI, the lexical entries for nouns define their PRED values, and either the specifier they combine with via the **head-spec** rule or the **bare-np** rule contributes the quantifier relation needed to bind the variable associated with the noun. Since this nominal affix will not go through either phrase structure rule needed for attaining a quantifier, and the lexical entry for the noun is not involved, this NI lexical rule needs to contribute both relations as EPs in the MRS.

In order to obtain the desired semantics, (i.e. a resulting MRS for NI sentences which is identical to their unincorporated paraphrases), I need to be able to connect the INDEX of the IN to a role in the verb's EP. I do this by identifying the ARG0 of the *noun-relation* contributed by this rule with the INDEX value of the first element on the verb's COMPS list (COMPS.FIRST). Because the IN will not combine with the verb via a phrase structure rule that targets the verb's COMPS list, the identification of arguments in this lexical rule is necessary for maintaining the desired semantics where the IN is the ARG2 of the verb's EP in the resulting MRS.

This rule on its own is only a partial analysis of NI. It will need to be paired with a second lexical rule which determines whether or not another noun is promoted to the direct object position of the verb or some material is doubled or stranded. Each of these non-inflecting lexical rules are defined in the following sections.<sup>6</sup>

For cases where material related to the IN appears outside of the verb (i.e. doubling or stranding), these phrases will go through phrase structure rules that allows them to combine with the verb syntactically in a way to interact with the IN semantically. These phrase structure rules will be introduced in section 4.2.4 below.

---

<sup>6</sup>Each of these secondary lexical rules inherits from a common supertype which specifies that the SPR and SPEC lists are empty which ensures that the verb does not combine with a specifier or act as a specifier itself.

### 4.2.2 Reduction noun incorporation lexical rules

In the case where noun incorporation does not lead to promotion, modifier stranding, or noun doubling, the verb will go through one of two reduction type lexical rules. The first of which applies to constructions where incorporation appears to change the transitivity of the verb, making an underlyingly transitive verb behave intransitively. The second of which applies to constructions where reduction incorporation does not appear to do so, and transitive verbs retain their transitive behavior (aside from taking an NP complement). Evidence for an impact on transitivity might come from (in)transitive case patterns, agreement morphology, etc. Based on my review of the literature, I am not aware of a language which employs both strategies. Therefore, these two lexical rules are mutually exclusive—a language with one rule in its grammar should not have the other.<sup>7</sup> Despite their differences, both rules function to prohibit the verb from syntactically combining with a direct object via the **head-comp** rule.

#### Resulting intransitive verbs

A lexical rule for modeling intransitive Reduction NI is shown in (60) below.

$$(60) \left[ \begin{array}{l} \textit{reduce-intrans-lex-rule} \\ \\ \text{SYNSEM ... VAL} \left[ \begin{array}{l} \text{SUBJ} \left\langle \left[ \begin{array}{l} \text{LOCAL} \left[ \begin{array}{l} \text{CONT.HOOK } \boxed{1} \\ \text{CAT} \left[ \begin{array}{l} \text{NCORP-MOD } \boxed{2} \\ \text{VAL} \\ \text{HEAD} \end{array} \right] \begin{array}{l} \boxed{3} \\ +np \end{array} \end{array} \right] \right] \right] \right\rangle \\ \text{COMPS } \boxed{4} \end{array} \right] \\ \\ \text{DTR.SYNSEM ...} \left[ \begin{array}{l} \text{SUBJ} \left\langle \left[ \begin{array}{l} \text{LOCAL} \left[ \begin{array}{l} \text{CONT.HOOK } \boxed{1} \\ \text{CAT} \left[ \begin{array}{l} \text{NCORP-MOD } \boxed{2} \\ \text{VAL} \end{array} \right] \begin{array}{l} \boxed{3} \\ \end{array} \end{array} \right] \right] \right] \right\rangle \\ \text{COMPS.FIRST } \boxed{4} \end{array} \right] \end{array} \right] \end{array} \right]$$

<sup>7</sup>As implemented, the customization system will only output a rule for one type or the other (see the discussion of implementation in chapter 5).

This rule accomplishes two main tasks. First, some, but not all, of the values of the element on the verb’s SUBJ list are identified between the daughter and the mother. HOOK — which includes the INDEX of this element — along with the subject’s valence and NCORP-MOD value are constrained to be the same on mother and daughter. NCORP-MOD is a new CAT feature that I define in section 4.2.4 for handling external modifiers of the IN. Instead of identifying the HEAD value of the subject between mother and daughter, I constrain the part of speech to be *+np* (noun or adposition), but allow the subject’s CASE value to be different on the mother than it is on the daughter. Thus if the lexical entry for a verb specifies that its subject must be marked with ergative case, this lexical rule provides a gap for which the CASE value on the verb’s SUBJ list may be modified.<sup>8</sup>

Second, the rule states identifies the value of the daughter’s COMPS.REST list with the mother’s COMPS list. If there is only one element on the daughter’s COMPS list, this will in effect ‘empty’ the mother’s. An empty COMPS list signals that the verb is not looking to combine with a complement in the syntax. Thus, what started as a transitive verb in the lexicon will behave intransitively as a result of this lexical rule.

### *Resulting Transitive Verbs*

For languages when Reduction NI results in a sentence with transitive behavior, a different lexical rule for Reduction NI is needed. This rule is presented in (61) below.

$$(61) \left[ \begin{array}{l} \textit{reduce-trans-lex-rule} \\ \text{SYNSEM.LOCAL.CAT.VAL} \\ \text{DTR.SYNSEM.LOCAT.CAT.VAL} \end{array} \left[ \begin{array}{l} \text{SUBJ} \quad \boxed{1} \\ \text{COMPS} \quad \boxed{2} \langle [\text{OPT} \quad +], \dots \rangle \\ \text{SUBJ} \quad \boxed{1} \\ \text{COMPS} \quad \boxed{2} \end{array} \right] \right]$$

---

<sup>8</sup>The updated case value would need to be defined by the user for the specific language. This step is described in chapter 5 on implementation.

In this rule, the subject is identified between daughter and mother,<sup>9</sup> and the COMPS list is not emptied. Instead, it is identified with that of its daughter, but constrained to have a positive boolean value for the OPT feature. This means that the element on the COMPS list is forced to be unexpressed orthographically, and the verb may not combine with another phrase via the `head-comp` rule in the syntax. Instead, an unary phrase structure rule for optional complements is applied to the verb to satisfy the COMPS list.

In this way, the verb remains syntactically transitive. If verbal agreement markers for objects are present such as in example (48) in Southern Tiwa, they will agree with the element on the COMPS list which is identified with the noun relation contributed by the incorporated noun affix.<sup>10</sup>

One effect of representing the IN as an unexpressed complement with the feature-value pair [OPT –] is that an additional ICONS *non-focus* element whose RARG is identified with the ARG0 of the IN is introduced in the MRS as seen in (62) (compare to (50)).

---

<sup>9</sup>There is no need to change the case value if incorporation has no impact on the behavior of the rest of the sentence.

<sup>10</sup>Specifying features on the incorporated noun is discussed further in chapter 5 on implementation.

$$(62) \left[ \begin{array}{l} mrs \\ LTOP \quad \boxed{h1} \ h \\ INDEX \quad \boxed{e2} \ e \\ \\ RELS \quad \left\langle \begin{array}{l} \left[ \begin{array}{l} \_see\_v\_rel \\ LBL \quad \boxed{h3} \ h \\ ARG0 \quad \boxed{e2} \\ ARG1 \quad \boxed{x4} \ x \\ ARG2 \quad \boxed{x5} \ x \end{array} \right] , \left[ \begin{array}{l} \_man\_n\_rel \\ LBL \quad \boxed{h6} \ h \\ ARG0 \quad \boxed{x5} \end{array} \right] , \left[ \begin{array}{l} exist\_q\_rel \\ LBL \quad \boxed{h7} \ h \\ ARG0 \quad \boxed{x5} \\ RSTR \quad \boxed{h9} \ h \\ BODY \quad \boxed{h8} \ h \end{array} \right] \end{array} \right\rangle \\ \\ HCONS \quad \left\langle \begin{array}{l} qeq \\ HARG \quad \boxed{h9} \\ LARG \quad \boxed{h6} \end{array} \right\rangle \\ \\ ICONS \quad \left\langle \begin{array}{l} non-focus \\ LARG \quad \boxed{e2} \\ RARG \quad \boxed{x5} \end{array} \right\rangle , \left\langle \begin{array}{l} non-focus \\ LARG \quad \boxed{e2} \\ RARG \quad \boxed{x4} \end{array} \right\rangle \end{array} \right]$$

*non-focus* elements are a way of representing information structure in a language. Song (2017) argues that argument optionality is relevant in expressing information structure in that a dropped element cannot be the focus of a sentence. Song (p.c.) suggests that NI might also have to do with information structure, reducing the focus on the IN. A full analysis of the impact of NI on information structure is worthwhile but falls beyond the scope of this thesis. For the time being, the discrepancy between the MRSs in (50) and (62) is not incompatible within my proposed analysis.

#### 4.2.3 Promotion noun incorporation lexical rules

When incorporation results in another argument syntactically taking the place of the verb's direct object, the verb will go through a Promotion NI lexical rule. In this section, I describe two such lexical rules that result in semantic representations for the promoted argument either being the possessor of the IN or an oblique argument of the verb. Unlike the two lexical rules for Reduction NI types, these two rules are not mutually exclusive in a language's grammar. A language might allow possessor raising and oblique argument promotion, though

not on the same token. The two lexical rules are structurally quite similar and only differ with respect to how indices are identified in order to attain the desired semantic representation.

### *Possessor Raising*

The lexical rule for possessor raising is given in (63) below. This rule accomplishes three main tasks. First, it identifies the subject of the daughter with that of the mother. Second, it specifies that the INDEX of the element on the COMPS list is different from that on the daughter—the IN. Third, it contributes a relation to the semantics and provides links between the semantic features of that predication and properties of the IN ([4] and [5]) as well as the possessor ([3]).

$$(63) \left[ \begin{array}{l} \textit{promote-poss-lex-rule} \\ \text{SYNSEM ... VAL} \left[ \begin{array}{l} \text{SUBJ [1]} \\ \text{COMPS} \left\langle \begin{array}{l} \text{OPT -} \\ \text{LOCAL} \left[ \begin{array}{l} \text{CAT} \left[ \begin{array}{l} \text{NCORP-MOD [2]} \\ \text{VAL.SPR } \langle \rangle \\ \text{HEAD } \textit{noun} \end{array} \right] \\ \text{CONT.HOOK.INDEX [3]} \end{array} \right] \end{array} \right\rangle, [6] \end{array} \right] \\ \text{DTR.SYNSEM ... VAL} \left[ \begin{array}{l} \text{SUBJ [1]} \\ \text{COMPS} \left\langle \begin{array}{l} \text{LOCAL} \left[ \begin{array}{l} \text{CAT.NCORP-MOD [2]} \\ \text{CONT.HOOK} \left[ \begin{array}{l} \text{INDEX [4]} \\ \text{LTOP [5]} \end{array} \right] \end{array} \right] \end{array} \right\rangle, [6] \end{array} \right] \\ \text{C-CONT.RELS} \left\langle \begin{array}{l} \textit{arg12-ev-relation} \\ \text{PRED "poss\_rel"} \\ \text{LBL [5]} \\ \text{ARG1 [4]} \\ \text{ARG2 [3]} \end{array} \right\rangle \end{array} \right] \end{array} \right]$$

By not identifying the INDEX values of the elements of the mother's and daughter's COMPS lists, this rule states that the verb is still seeking to combine syntactically with some noun, and ensures that the noun is not co-indexed with the IN. The CAT values of the element

on the COMPS list ensure that the complement is a noun phrase by specifying that its HEAD value is *noun* and that it has an empty SPR list. In addition to specifying different INDEX values of the element on the mother and daughter's COMPS lists, this rule does not identify most features between these two elements. The only feature identified between the promoted object and the IN is the NCORP-MOD feature. This is a novel feature for external NI modification described in the following section, but it is essential that the promoted object has a negative NCORP-MOD value, and identification ensures this is the case. The [OPT –] feature on this new promoted object ensures that it is overt, as we would not expect an element to be promoted, just to be dropped due to argument optionality in the language.

To obtain the correct possession semantics, the ARG1 of this relation is identified with the index of the possessum (the IN), and the LBL of the *poss\_rel* is also identified with the possessum's LBL via the LTOP feature. The ARG2 of the relation is identified with the index of the possessor (in this case, the promoted object).

### *Oblique Argument Promotion*

The lexical rule for oblique argument promotion is shown in (64) below.

$$(64) \left[ \begin{array}{l} \text{promote-obl-lex-rule} \\ \text{SYNSEM} \left[ \begin{array}{l} \text{CONT.HOOK.LTOP } \boxed{7} \\ \dots \text{VAL} \left[ \begin{array}{l} \text{SUBJ } \boxed{1} \\ \text{COMPS} \left\langle \begin{array}{l} \text{OPT } - \\ \text{LOCAL} \left[ \begin{array}{l} \text{CAT} \left[ \begin{array}{l} \text{NCORP-MOD } \boxed{6} \\ \text{VAL.SPR } \langle \rangle \\ \text{HEAD } \boxed{4} \textit{noun} \end{array} \right] \\ \text{CONT.HOOK.INDEX } \boxed{3} \end{array} \right] \end{array} \right] \end{array} \right] \end{array} \right] \\ \text{DTR ... LOCAL} \left[ \begin{array}{l} \text{SUBJ } \boxed{1} \\ \text{CAT.VAL} \left[ \begin{array}{l} \text{COMPS} \left\langle \begin{array}{l} \text{LOCAL.CAT} \left[ \begin{array}{l} \text{NCORP-MOD } \boxed{6} \\ \text{HEAD } \boxed{4} \end{array} \right] \end{array} \right] \end{array} \right] \\ \text{CONT.HOOK.INDEX } \boxed{2} \end{array} \right] \\ \text{C-CONT.RELS} \left\langle \begin{array}{l} \textit{arg12-ev-relation} \\ \text{LBL} \quad \quad \quad \boxed{7} \\ \text{ARG1} \quad \quad \quad \boxed{2} \\ \text{ARG2} \quad \quad \quad \boxed{3} \end{array} \right\rangle \end{array} \right] \end{array} \right]$$

This rule, like that in (63) above, identifies the entire SYNSEM value of the subject between mother and daughter, but not that of the complement. It similarly constrains the new complement to be [NCORP-MOD –] by identifying this value between mother and daughter.<sup>11</sup>

The HEAD value of the element on the COMPS list is identified between mother and daughter. This allows the characteristics of the promoted object (such as the part of speech) to behave as intended by the verb's lexical entry. In my review of the literature, promoted objects always have the same part of speech and case value as unincorporated objects, but if a language is identified where this is not the case, the library would need to be updated to account for this.

This rule differs from that above with respect to the C-CONT which handles the semantic contributions. Here, a relation is defined whose ARG1 value is identified with the

---

<sup>11</sup>The NCORP-MOD value is constrained to be negative on the daughter by the definition of lexical items which will be explained in the implementation chapter (§5).

INDEX value of the verb itself, and the ARG2 is identified with the INDEX of the promoted argument. The design for this semantic relation comes from the handling of object-addition operations in the Grammar Matrix’s valence change library (Curtis, 2018). Instead of identifying the added object’s index to the list of ARG $n$  in the verb relation, a separate relation is added to the MRS. The ARG1 of this new relation is identified with the verb and the ARG2 with the added object. In the rule above, the PRED value for the added relation is left underspecified as to accommodate the promotion of different argument types (i.e. locative, benefactive, etc.). Structurally, all would behave the same way. As will be explained in the following chapter, the analysis treats the predicate value of the EP which relates the promoted argument and the verb as user-defined in the customization system. For the sentence in (53), the user would have designated the value to be `benefactive_rel`.

#### 4.2.4 *Modifier Noun Incorporation*

In this section, I analyze NI strategies in which some material which is semantically related to the IN appears outside of the verb complex. This involves two related but distinct constructions—doubling and stranding. In doubling, a co-referential, and usually a more specific version of the IN, appears in object position. In stranding, some modifier of the IN appears outside the verb, such as an adjective or demonstrative. Both constructions require separate analyses, but overlap in the presence of some external material which is linked in some way to the IN. In this section, I define two lexical rules for incorporating verbs with doubling and stranding, and two unary phrase-structure rules for doubled and stranded material.

In order to constrain these elements, and only these elements, to combine with a verb having undergone incorporation, I define a new boolean feature NCORP-MOD (short for noun incorporation modifier) which is a feature inside of CAT. NCORP-MOD will be defined in a grammar for any language which has either doubling, stranding, or both. The value for this feature will be negative almost everywhere in the grammar, but positive for elements which can modify an incorporated noun. The positive value on the modifier or doubled noun is assigned by one of two non-branching phrase structure rules defined in (66) and

(68) in the following sections. A lexical rule will apply to the verb which specifies that the NCORP-MOD value must be positive on the element on the verb's COMPS list. Thus, the verb with an IN seeking a doubled noun or stranded modifier can syntactically combine with appropriate material after the application of both the lexical and phrase structure rules.

### *Doubling*

The lexical rule for doubling is shown in (65).

$$(65) \left[ \begin{array}{l} \textit{double-lex-rule} \\ \text{SYNSEM ... VAL} \left[ \begin{array}{l} \text{SUBJ} \quad \boxed{1} \\ \text{COMPS} \left\langle \begin{array}{l} \text{OPT -} \\ \text{LOCAL} \left[ \begin{array}{l} \text{CAT} \left[ \begin{array}{l} \text{NCORP-MOD +} \\ \text{HEAD } \textit{noun} \end{array} \right] \\ \text{CONT.HOOK} \left[ \begin{array}{l} \text{LTOP} \quad \boxed{2} \\ \text{INDEX} \quad \boxed{3} \end{array} \right] \end{array} \right] \end{array} \right] \end{array} \right] \end{array} \right] \\ \text{DTR ... VAL} \left[ \begin{array}{l} \text{SUBJ} \quad \boxed{1} \\ \text{COMPS} \left\langle \left[ \text{LOCAL.CONT.HOOK} \left[ \begin{array}{l} \text{LTOP} \quad \boxed{2} \\ \text{INDEX} \quad \boxed{3} \end{array} \right] \right] \right\rangle, \boxed{4} \end{array} \right] \\ \text{C-CONT} \left[ \begin{array}{l} \text{RELS} \quad \langle \rangle \\ \text{HCONS} \quad \langle \rangle \\ \text{ICONS} \quad \langle \rangle \end{array} \right] \end{array} \right]$$

In this lexical rule, the SUBJ value is identified between the mother and the daughter. The first element on the mother's COMPS list is constrained to have a positive NCORP-MOD value. This specification ensures that when the verb combines with a complement via the head-comp rule, that complement has a positive NCORP-MOD value. As mentioned, the only way for phrases to have a positive NCORP-MOD value is to go through an associated phrase structure rule.

This rule also identifies the LTOP and INDEX values of the daughter's COMPS.FIRST

element with that of the mother. Recall that because of the general NI lexical rule in (59), the first element on the daughter's COMPS list is the IN itself. This identification above does not occur because we expect to pick up the IN via the **head-comp** rule, but rather is necessary to expose the LTOP and INDEX values of the IN when the doubled noun or stranded modifier is picked up by the **head-comp** rule. In order to obtain the desired semantic representation for the clause, these values need to be accessible for the proper argument identification. At this point, the analysis presents an unusual mapping between syntax and semantics, one that is not anticipated by Copestake et al. (2001).

The more specific noun (in example (55), *rabahbót* 'bullhead') must go through the unary phrase structure rule in (66) in order to combine with the verb and IN in the syntax. This rule accomplishes two main tasks: it sets the NCORP-MOD value of the constituent to positive, and creates a new semantic relation to model the relationship of the incorporated and doubled nouns.



of the IN's index is made possible by the C-CONT.HOOK values. The values of the C-CONT.HOOK LTOP and INDEX will map to the IN once this phrase has combined with the verb via the *head-comp*. The NCORP-MOD value specifications make this unification valid, and the fact that the incorporated noun's INDEX and LTOP values were preserved in the lexical rule in (65) allow for them to be accessed as the C-CONT.HOOK INDEX and LTOP values in this rule. Finally, the LBL of the new relation is identified with the IN's LTOP because this relation should have the same scope as the IN.

### *Stranding*

The lexical rule for stranding is shown in (67).

$$(67) \left[ \begin{array}{l} \textit{strand-lex-rule} \\ \\ \text{SYNSEM ... VAL} \\ \\ \text{DTR ... VAL} \\ \\ \text{C-CONT} \begin{array}{l} \text{RELS} \quad \langle \rangle \\ \text{HCONS} \quad \langle \rangle \\ \text{ICONS} \quad \langle \rangle \end{array} \end{array} \right] \left[ \begin{array}{l} \text{SUBJ} \left\langle \begin{array}{l} \text{LOCAL} \left[ \begin{array}{l} \text{CONT.HOOK } \boxed{1} \\ \text{CAT} \left[ \begin{array}{l} \text{NCORP-MOD } \boxed{2} \\ \text{VAL } \boxed{2} \\ \text{HEAD } +np \end{array} \right] \end{array} \right] \right\rangle \\ \\ \text{COMPS} \left\langle \begin{array}{l} \text{OPT } - \\ \text{LOCAL} \left[ \begin{array}{l} \text{CAT} \left[ \begin{array}{l} \text{NCORP-MOD } + \\ \text{HEAD } +jp \end{array} \right] \\ \text{CONT.HOOK} \left[ \begin{array}{l} \text{LTOP } \boxed{5} \\ \text{INDEX } \boxed{6} \end{array} \right] \end{array} \right] \end{array} \right\rangle, \boxed{4} \right\rangle \\ \\ \text{SUBJ} \left\langle \begin{array}{l} \text{LOCAL} \left[ \text{CONT.HOOK } \boxed{1} \right] \\ \text{CAT} \left[ \begin{array}{l} \text{NCORP-MOD } \boxed{2} \\ \text{VAL } \boxed{3} \end{array} \right] \end{array} \right\rangle \\ \\ \text{COMPS} \left\langle \begin{array}{l} \text{LOCAL.CONT.HOOK} \left[ \begin{array}{l} \text{LTOP } \boxed{5} \\ \text{INDEX } \boxed{6} \end{array} \right] \end{array} \right\rangle, \boxed{4} \right\rangle \end{array} \right]$$

This lexical rule is very similar to that in (65) above, but differs in the identification of

subject information between mother and daughter, and the HEAD value on the first element of the mother's COMPS list. This rule does not identify the entire SUBJ list between the mother and daughter, instead it identifies the HOOK, NCORP-MOD, and VAL features like the lexical rule for intransitive Reduction NI in (60). This allows for the CASE value on the mother to be different than that on the daughter. The first element on the mother's COMPS list is constrained to have a HEAD value of *+jp*. This is a disjunctive type used in the Grammar Matrix that includes both adjective and adpositions which can both be nominal modifiers.

The non-branching phrase structure rule for stranded modifiers is provided in (68) below.

$$(68) \left[ \begin{array}{l} \text{adj-ni-mod-phrase} \\ \\ \\ \end{array} \right] \left[ \begin{array}{l} \text{SYNSEM.LOCAL.CAT} \\ \\ \text{HEAD-DTR ... LOCAL} \\ \\ \text{C-CONT} \end{array} \right] \left[ \begin{array}{l} \text{VAL} \\ \\ \text{CAT} \\ \\ \text{RELS} \langle \rangle \\ \text{ICONS} \langle \rangle \\ \text{HCONS} \langle \rangle \\ \text{HOOK} \end{array} \right] \left[ \begin{array}{l} \left[ \begin{array}{l} \text{SUBJ} \quad \boxed{1} \\ \text{COMPS} \quad \boxed{2} \\ \text{SPEC} \quad \boxed{3} \\ \text{SPR} \quad \langle \rangle \end{array} \right] \\ \\ \left[ \begin{array}{l} \text{HEAD} \left[ \begin{array}{l} \text{adj} \\ \text{MOD.FIRST...INDEX} \quad \boxed{4} \end{array} \right] \\ \text{VAL} \left[ \begin{array}{l} \text{SUBJ} \quad \boxed{1} \\ \text{COMPS} \quad \boxed{2} \\ \text{SPEC} \quad \boxed{3} \end{array} \right] \\ \text{NCORP-MOD} - \\ \text{CONT.HOOK.LTOP} \quad \boxed{5} \end{array} \right] \\ \\ \left[ \begin{array}{l} \text{LTOP} \quad \boxed{5} \\ \text{INDEX} \quad \boxed{4} \end{array} \right] \end{array} \right]$$

Like in the rule for doubled nouns in (66), this non-branching phrase structure rule specifies that the NCORP-MOD value is negative on the daughter and positive on the mother,

allowing the syntax to correctly combine this phrase with a verb and incorporated noun and prohibiting the rule from feeding itself. It differs from the doubling rule in how arguments are identified in order to obtain the desired semantic representation. Here, the INDEX associated with the IN is identified with the INDEX value on the modifier’s MOD list, and their LTOP values are also identified. This identification is consistent with how adjectives are analyzed elsewhere in Matrix grammars (Trimble, 2014).

Unlike (66), this rule does not contribute an EP relation into the semantics. Instead, the INDEX of the IN, accessed via the C-CONT.HOOK feature, is identified with the element on the stranded modifier’s MOD list. Almost all of the VAL lists are identified between mother and daughter, except the SPR list is constrained to be empty on the mother, ensuring that this stranded modifier does not combine syntactically with a specifier.

### 4.3 Summary

In this chapter, I described my analysis of the different variations of noun incorporation within the frameworks of HPSG and MRS. A verb incorporating a noun will always go through a common lexical rule which affixes the noun to the verb, and contributes the noun relation and associated quantifier relation to the semantics. The verb will then go through a second lexical rule depending on the behavior of the rest of the clause. For Reduction NI, the lexical rule will either clear the verb’s COMPS list or specify that the complement must be unexpressed (depending on whether or not the resulting verb behaves transitively or intransitively). For Promotion NI, the lexical rule re-assigns the INDEX value on the verb’s COMPS list to allow the verb to combine with a new, promoted object in the syntax. Two lexical rules of this type are described which model different semantic representations for this promoted object—one in which it is the possessor of the IN, and one in which it is an oblique argument. For Modifier NI, I define a new boolean feature, NCORP-MOD, which restricts what material may combine with and modify the IN of a verb. Two lexical rules for this type are described which require the verb’s complement to be [NCORP-MOD +]. Two corresponding non-branching phrase structure rules create phrases which are positive

for this feature and create the desired semantic relations which model the relationship of the doubled or stranded material and the incorporated noun. In the next chapter, I will describe how I implemented this analysis into the Grammar Matrix customization system, with targeted elicitation questions on the user-facing webpage to determine which parts of my analysis are relevant for a given language.

## Chapter 5

### IMPLEMENTATION

This chapter discusses how I implemented my syntactic and semantic analysis of Noun Incorporation into the Grammar Matrix. This process involved designing a system of elicitation questions in a user-facing web questionnaire. The answers to users' questions will determine which aspects of my analysis are needed to include in their grammar to properly account for the variation of NI in their language of choice. This chapter describes the three main tasks involved in implementation: the questionnaire design (5.2), backend programming to map users' answers to TDL (Copestake, 2002a) in their grammar files (5.3), and validation testing to ensure that users make compatible choices in the customization system (5.4).

#### **5.1 Background**

Users wanting to use the Grammar Matrix as a starting point for building an implemented grammar begin by visiting the customization system (Bender et al., 2002, 2010).<sup>1</sup> The various subpages target different linguistic phenomena which may or may not be present in the language. The user is prompted to answer questions in the form of radio buttons, check boxes, drop-down menus, or text fields. When they save their choices on a particular page, validation tests run to ensure that the combination of choices is valid. These validations come in the form of either errors or warnings, designated as red asterisks or question marks respectively on the webpage in the area of concern. Errors will prohibit the combination of choices from generating a functional grammar and must be addressed before downloading the grammar. Warnings won't prevent a grammar from being downloaded, but are meant to catch combinations of choices that might produce results that the user did not expect or

---

<sup>1</sup><https://matrix.ling.washington.edu/customize/matrix.cgi>

that lead to grammars that will not compile in DELPH-IN processors such as the Linguistic Knowledge Builder (LKB) (Copestake, 2002b) or Answer Constraint Engine (ACE) (Crysmann and Packard, 2012). For instance, some choices are conditional upon others being made first, and will be ignored by the Matrix code if the others are not. Validation warnings are meant to guide the user to make complete choices.

Once the user is happy with their choices and no validation errors exist, the user is able to download the grammar. Validation warnings will not prevent downloads, but should ideally be addressed prior to downloading the grammar. At this step, the backend Matrix code takes in the user’s choices in the form of a text file known as the choices file, and uses this input to write the necessary lexical entries, lexical rules, phrase structure rules, and other relevant types to the correct grammar files. These files work in conjunction with the core grammar (found in `matrix.tdl`) to parse and generate strings in the language. The files are written in Type Description Language (TDL), which is a formalism for representing the typed feature structures of HPSG (Copestake, 2002a). In this customization step, the various libraries which make up the Grammar Matrix are called one by one, each interacting with the information in the choices file, and contributing relevant analyses to the grammar files.

## **5.2 Questionnaire Design**

My contributions to the customization system to elicit the grammatical behavior of NI from users exist in two locations of the webpage. I created a new subpage titled ‘Noun Incorporation’ where users select from which types of NI exist in the language they are describing. I also extend the morphology subpage. Here, users can create a position class for incorporated nouns which contribute a semantic predicate to the MRS. The creation of a new subpage and extension of the morphotactics subpage are described in 5.2.1 and 5.2.2 respectively.

### *5.2.1 Noun Incorporation Subpage*

When visiting the Noun Incorporation subpage in the customization system, the user will first see a brief description of what NI is, and a warning that this page will only elicit

## Noun Incorporation [\[documentation\]](#)

Noun Incorporation is the process of morphologically incorporating a noun, usually the direct object of the verb, into the verb itself resulting in a complex verb.

Noun Incorporation occurs in this language.

Figure 5.1: First checkbox on NI subpage indicating that NI occurs in the language

information regarding the incorporation of a verb's direct object.<sup>2</sup> If the user would like to make use of my analysis, the first thing that will need to be done is selecting the check box which states that NI occurs in the language. Selecting this check box will enable the user to create a position class for the incorporated noun which will be described in the following section. The description and checkbox are shown in Figure 5.1.

The rest of the page is broken into three main parts which correspond to my analysis of three main types of NI: Reduction NI, Promotion NI, and Doubling/Stranding NI. These three sections each have a check box associated with them and a description of what that type of NI looks like in a language. Each are presented and described below. If the user has stated that NI occurs in the language, they may select one, two, or all of the three types presented here. For each type, additional information is needed from the user regarding the exact behavior of that type.

### *Reduction NI*

In NI constructions where there is no other overt expression of the direct object argument, the user is prompted to select the checkbox associated with Reduction NI. If checked, the user is then prompted to indicate whether reduction NI results in a clause that appears to be

---

<sup>2</sup>As introduced in the literature review and addressed in the analysis, I do not provide functionality for incorporating subjects or oblique arguments. This is left to future work.

The direct object of the verb incorporates and there is no other overt expression of this argument.

If so, how does the verb behave?

- The verb appears transitive (i.e. transitive morphology, case)
- The verb appears intransitive (i.e. intransitive morphology, case)

If intransitive, does the case marked on the subject change?

- No
- Yes. New case:  ▼

Are there any position classes that this type of NI should not co-occur with?

Figure 5.2: Choices capturing the behavior of Reduction NI

transitive or intransitive. These two options are marked with radio buttons given that only one strategy should be appropriate for a language. This section of the subpage is shown in Figure 5.2.

Further, in the case where resulting sentences appear intransitively, the user is given the option to define a new case value for the subject. For example, in the case of a language with ergative case marking on transitive subjects and absolutive case marking on intransitive subjects, the subject of a verb with NI should be absolutive. An illustration of selecting these choices and defining a different case value as a result of incorporation is shown in Figure 5.3.

The final part of the Reduction NI section asks the user if there are any position classes that this type of NI should not occur with. This is known as a forbid constraint, and is usually definable on the morphology subpage. The Grammar Matrix handles forbid constraints by creating flags for the types. For example, if Reduction NI in a language caused the resulting sentence to behave intransitively and the language also has both transitive and intransitive agreement morphology, the user would be able to forbid Reduction NI construc-

The direct object of the verb incorporates and there is no other overt expression of this argument.

If so, how does the verb behave?

- The verb appears transitive (i.e. transitive morphology, case)  
 The verb appears intransitive (i.e. intransitive morphology, case)

If intransitive, does the case marked on the subject change?

No

Yes. New case:  ▼

Are there any position classes  ergative  absolutive of NI should not co-occur with?

Add a Forbid constraint

Figure 5.3: Example user input for a language where Reduction NI results in an intransitive sentence with a different case value on the subject.

tions from occurring with transitive morphology. This is important because each type of NI defined on this subpage will be represented as an LRT in the same position class (see section 5.3.1). Therefore, if both Reduction and Promotion NI occur in the language, and both should combine with different agreement morphemes, this position class will be input to both. These forbid constraints allows the user to ensure that the correct type of NI co-occurs with the correct agreement morphology.

### *Promotion NI*

The second chunk on this subpage allows the user to define a strategy for Promotion NI. The description asks if another argument (ex: possessor, locative, instrumental) becomes the new direct object of the verb. If this box is selected, the user must then indicate what type of argument can be promoted. There are two options here listed, and the user is able

The direct object of the verb incorporates and another argument (such as a possessor, locative, instrumental, etc.) becomes the new direct object.

If so, what type of noun is promoted to object position?

The possessor of the incorporated noun is raised to object position

An oblique argument is promoted to object position. You may define a predicate name for this process. Predicate:

Are there any position classes that this type of NI should not co-occur with?

Figure 5.4: Choices capturing the behavior of Promotion NI

to select one or both options. This section is shown in Figure 5.4.

If possessor raising occurs in the language, the user simply selects the first option. If oblique arguments can be promoted, the user selects the second option, and is prompted to define a predicate value for the type of promoted argument such as ‘benefactive.’<sup>3</sup> This predicate value is stored as a string in the choices file and set as the PRED value of the promoted argument relation in the MRS. As of now, there is only the option to define one predicate value, though if more than one type of oblique argument promotion is available in the language, the user can simply duplicate this lexical rule in TDL editing and change the associated PRED value on the second rule, or enter a predicate value that is underspecified between more than one. The first option would lead to parsing ambiguity, but would ensure that the correct semantic representation for each sentence is achieved. The EP structure for each type would be identical, therefore I provide the functionality for one oblique promotion lexical rule type, and invite the user to duplicate as necessary for their language.

---

<sup>3</sup>Not providing a predicate value leads to a validation error which will prohibit the user from being able to download the grammar.

Elements of the incorporated noun appear outside of the verb. For example, a co-referent noun that is more specific than the incorporated noun, modifiers or demonstratives of the incorporated noun, or a copy of the incorporated noun itself.

If so, what elements can be stranded?

A more specific noun or doubled noun

Are there any position classes that this type of NI should not co-occur with?

Add a Forbid constraint

A modifier (ex: adjective, demonstrative, quantifier)

Are there any position classes that this type of NI should not co-occur with?

Add a Forbid constraint

Figure 5.5: Choices capturing the presence or absence of Doubling and Stranding NI

### *Doubling/Stranding NI*

For doubling and stranding, both strategies are grouped in the third section of subpage questions and separated into two different check boxes as shown in Figure 5.5. As for Promotion NI strategies, a user is able to select one or both options here as both are allowed within a single grammar. Each appears with the opportunity to define separate forbid constraints.

### *Summary*

This section walked through the noun incorporation subpage on the customization system. On this subpage, the user is presented with questions which help them identify which types of NI exist in the language they are modeling. The subpage is organized around the three types of NI analyzed: Reduction NI, Promotion NI, and Modifier NI. Radio buttons allow

the user to choose either the option for transitive or intransitive Reduction NI as only one should be applicable to a given language. The other types are selectable with check boxes, which allow the user to choose any combination of Promotion NI types (possessor raising and oblique argument promotion) and Modifier NI types (doubling and stranding). Any additional information needed from the user appears as a nested question within a type they have chosen.

### 5.2.2 *Morphology Subpage*

If the user has indicated that NI occurs in the language on the Noun Incorporation subpage, they will now be able to define a position class for incorporating nouns on the Morphology subpage. Under ‘Verb inflection,’ the user chose the button to add a position class. Here, they would proceed as normal for defining PCs, including defining a name (such as ‘noun incorporation’), specifying if it is a prefix or a suffix, choosing inputs, and defining any requires or forbids constraints as necessary.

Then, instead of adding a lexical rule type, the user would add an incorporated stem lexical rule type (see Figure 5.6). This follows the same basic mechanism as incorporated adjectives for noun inflection (Trimble 2014). Similarly to normal LRTs, the user will define the name, optional supertypes, and morphotactics constraints for the IS-LRT. In addition, they will also define the orthography and predicate values of the IN. I have also included the ability to define features to be specified on the IN. This is important for languages with transitive type reduction NI where agreement morphology on the verb agrees with the IN. An example of a fully-defined IS-LRT is shown in Figure 5.7.

If an noun in the language can appear in argument position as well as affixed to the verb as an IN, the user will need to define it twice in the customization system: both in the lexicon and as an IS-LRT in the morphology.<sup>4</sup> However, this duplication comes with an

---

<sup>4</sup>Future work on this library could automate this process. For example, if there is a type of noun which incorporates in the language, the user could define a noun class for this type, and the system could populate the orthography and predicate values for IS-LRTs with those from each noun in this class.

## Verb Inflection

### ▼ noun\_incorporation (verb-pc1)

X **Verb 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:

▶ verb-pc1\_lrt1

Figure 5.6: Once NI is defined in the language, the user may now define Incorporated Stem Lexical Rule Types.

▼ incorporated\_noun (verb-pc1\_is-lrt1)

**Incorporated Stem Lexical Rule Type 1:**

Name:

Supertypes:  ▼

Features:

Name:  Value:  ▼ Specified on:  ▼

Morphotactic Constraints:

Lexical Rule Instances:

Instance 1

No affix

Affix spelled  Predicate:

Figure 5.7: IS-LRT with features defined on the IN, orthography and predicate value defined.

added orthographic benefit. In some languages such as Chukchi [ckt] (Gerdt, 2017), the orthographic form of the noun when incorporated is different than when it appears outside of the verb. The user may then spell the noun differently in the two locations to account for this.

### *5.2.3 Summary*

In this section, I introduced my additions to the user-facing element of the customization system. I created a new subpage for Noun Incorporation which allows the user to indicate which types of NI are found in the language they are modeling. I also described additions made to the Morphology subpage. Once noun incorporation is enabled on the new subpage, users are able to define incorporating stem LRTs and LRIs within a verbal position class. Within the IS-LRT, users can define features marked on the IN, and within IS-LRIs, users can designate the spelling and predicate values of INs.

## **5.3 Python Code Implementation**

As outlined in the analysis chapter, all of my NI strategies involve at the minimum two lexical rules that a verb will go through. The first contributes the semantic information of the IN and the second does the necessary handling of the verb's COMPS list to correctly model the valence of the verb. In terms of customization, a lexical rule is represented by the user's definition of PCs and their nested LRTs. Since noun incorporation involves only the affixation of one morpheme (the IN), I wanted to design my library in a way where the user would only need to define one PC for noun incorporation. Within this PC, the user could define one or more LRTs as necessary<sup>5</sup> and as many IS-LRIs for each noun in the language that can be incorporated. In this section, I describe this single position class approach and how my code handles the second required lexical rule behind the scenes.

---

<sup>5</sup>For example, the user could define different IS-LRTs for different gender values marked on INs.

### 5.3.1 *A Single Position Class Approach*

When the user defines a PC with an IS-LRT, my code writes a new ‘ghost’ position class from scratch named ‘NI-valence’, a nod to the fact that any of the NI variations impact the VAL lists of the verb. The ‘ghost’ PC takes the user-defined NI PC as input and becomes the input to whatever the user specified their PC could be the input of. This approach takes some of the burden away from the user. They will not need to manually define non-inflecting lexical rules for each different type of NI in the language. Instead, they will define one PC for NI in general, corresponding to the first part of my analysis (see 4.2.1). The second part of my analysis is then taken care of behind the scenes, based on the user’s choices on the Noun Incorporation subpage.

For each of the NI strategies selected by the user on the Noun Incorporation subpage, I algorithmically write LRTs of the ‘ghost’ PC to the choices file. Each of these LRTs is constrained to be non-inflecting and is named after the type of NI (either ‘reduce’, ‘promote-poss’, ‘promote-obl’, ‘double-noun’, or ‘strand-mod’).<sup>6</sup> Any forbid constraint that the user defines for a NI type is associated with the corresponding LRT in the ‘ghost’ PC.

All of these additions are added to the choices file upon customization. They will not show up on the copy of the choices file included in the grammar folder because the step of adding the choices file to the output directory happens before this step. In other words, users will not see the effects of this backend process on their choices file, but will see the additional lexical rules types appear in their `mylang.tdl` file upon downloading their grammar. The rest of the customization process which takes the choices file as input will however have access to this new PC and LRTs, including functions which write flags or calculate rule supertypes.

### 5.3.2 *Implementing the three types of NI*

TDL corresponding to the analyses introduced in the previous chapter are stored in a python file and input directly to the user’s `mylang.tdl` file if they indicated that type of NI was

---

<sup>6</sup>There is only one name for reduce since the two choices on the page are mutually exclusive.

present in the language. I account for some minimal additional customization (ex: predicate values for promoted oblique arguments). In this section, I describe how this process happens, making use of the elicitation questions presented in the previous sections.

Over the course of development, I refined my analysis and implementation with the use of natural and pseudo languages. In one particular case, the language in question, Southern Tiwa [tix],<sup>7</sup> had instances of incorporation and passivization occurring in the same sentence. The argument incorporated into the verb was the argument demoted to object due to passivization. This prompted me to investigate the interaction of my implemented analysis alongside the existing valence change library (Curtis, 2018). I found a weakness in my analysis that I would not have considered without testing this interaction. Some valence change operations, particularly applicatives and causatives, add elements to the verb's COMPS list. Since my analysis was only making reference to the sole element on the verb's COMPS list, it was incompatible with a COMPS list with more than one element. I updated my analysis and the associated TDL so that each lexical rule handling NI will instead make specific reference to the first thing on the COMPS list which allows for COMPS lists of variable length.

The first step in customizing NI is done when my `noun_incorporation.py` file is called in the customization process. The choices file is modified with a new position class for NI valency and the various LRTs for the different types of NI indicated to be present in the language. After the choices file is edited, it is then processed by the morphotactics library (O'Hara, 2008, Goodman, 2013) when `morphotactics.py` is called.

`morphotactics.py` handles all tasks needed to configure the lexical rule types which define both inflecting and non-inflecting lexical rules in the grammar. The library processes any dependencies between PCs and LRTs with flags and intermediate types. Flags are used to ensure that lexical rules which must co-occur with other lexical rules do so, and keeps incompatible rules from applying in the same inflectional process. Intermediate types

---

<sup>7</sup>For more explanation, see (6.1.2) the evaluation chapter.

determine which rules can feed others. If a user indicated that one PC is input to another, it will inherit from the intermediate type which is compatible with the daughter of the other rule. An example of this can be seen in (69) from the Tongan [ton] grammar I made during development. Here, the type for transitive verb lexical items inherits from the type *NI-valence-rule-dtr*, which means that transitive verbs can unify with the DTR feature of the first noun incorporation lexical rule (i.e. the NI-valence lexical rule).

```
(69) transitive-verb-lex := transitive-lex-item & NI-valence-rule-dtr &
      [ SYNSEM.LOCAL.CAT.VAL.COMPS < #comps >,
        ARG-ST < [ LOCAL.CAT.HEAD adp ],
          #comps &
            [ LOCAL.CAT cat-sat &
              [ VAL [ SPR < >,
                COMPS < > ],
                HEAD adp ] ] > ]].
```

The ‘ghost’ position class and its non-inflecting LRTs contributed behind the scenes are processed in just the same way as those defined by the user. In the output grammar, the types for these rules will be elaborated by whatever flags and intermediate types necessary.

The morphotactics library also assigns the appropriate supertypes to PCs and LRTs. Lexical rule supertypes determine which values are identified and which can be different between mother and daughter in a lexical rule. For instance, *cat-change-only-lex-rule* is a supertype that allows a lexical rule to assign a different value to any CAT feature on the mother of the rule. It copies up all other values. *infl-lex-rule* is the supertype for all lexical rule which are inflecting, and *const-lex-rule* is the supertype for non-inflecting rules. For the valence NI ‘ghost’ PC, it will be assigned the *const-lex-rule* supertype since all of the associated LRTs are non-inflecting. The user-defined NI PC with IS-LRTs will inherit from *infl-lex-rule*, since the IN will orthographically appear affixed to the verb root. Since the ‘ghost’ PC impacts the valence lists of the verb and some of its LRTs add EPs to the semantics, this PC will also inherit from *add-ccont-val-change-only-lex-rule*. For those LRTs which don’t contribute to the semantics, they inherit from other types which constrains the C-CONT feature, or directly state that the lexical rule type does not contribute

a semantic relation with empty RELS, HCONS, and ICONS lists. This allows the lexical rules for the various types of NI to add semantics relations via the C-CONT.RELS list, and to add or remove elements from the verb's COMPS list (which is inside VAL).

Once the supertype and flag information is correctly stored for each PC and LRT, the code moves to writing the TDL to the necessary files: `mylang.tdl`, `lrules.tdl`, `irules.tdl`, and `rules.tdl`. First, the PCs then the LRTs are translated to TDL. Both will become lexical rule types defined in `mylang.tdl`, with the LRT types inheriting from their associated PC types. The lexical rules themselves will then be instantiated in either `lrules.tdl` or `irules.tdl`, depending on whether the rule is inflectional (`irules`) or not (`lrules`). When the LRTs are being written, my addition to the morphotactics library—`write_noun_incorporation_behavior()`—is called.

`write_noun_incorporation_behavior()` takes the choices file as input and matches the user's answers to the TDL output needed. For any language that is indicated to have NI, TDL corresponding to the first part of my analysis (i.e. the lexical rule in (59) introduced in the previous chapter) is added to `mylang.tdl` and is shown in (70).<sup>8</sup>

```
(70)  noun-incorp-lex-rule-super :=
      [ SYNSEM.LOCAL.CAT.VAL.COMPS [ FIRST.LOCAL.CONT.HOOK [ INDEX #ind,
                                                                LTOP #larg ],
                                                                REST #rest ],
        DTR.SYNSEM.LOCAL.CAT.VAL.COMPS.REST #rest,
        C-CONT [ HCONS.LIST < qeq &
                  [ HARG #harg,
                    LARG #larg ] >,
                ICONS.LIST < >,
                RELS.LIST < noun-relation &
                  [ ARGO #ind,
                    LBL #larg ],
                  quant-relation &
                  [ PRED "exist_q_rel",
                    ARGO #ind,
                    RSTR #harg ] > ] ].
```

---

<sup>8</sup>The name of this lexical rule type will depend on what the user defines as the name of the PC for noun incorporation, with `-lex-rule-super` appended to the end.

TDL corresponding to the NI Valence PC is then added to `mylang.tdl`, and is shown in (71)<sup>9</sup> below. As a PC, this TDL will be a supertype that its LRTs inherit from.

```
(71)  ni-valence-lex-rule :=
      [ SYNSEM.LOCAL.CAT.VAL [ SPEC < >,
                              SPR < > ] ].
```

### *Reduce lexical rule*

If the LRT name is ‘reduce’, my code finds the value associated with the key ‘red-val’ in the choices file. This value may either be ‘red-trans’ for transitive Reduction NI behavior or ‘red-intrans’ for intransitive. If it is intransitive, the following TDL is added to `mylang.tdl`. This TDL corresponds to my analysis of intransitive Reduction NI shown in the lexical rule in (60) in the previous chapter.

```
(72)  reduce-lex-rule := val-change-only-lex-rule &
      [ SYNSEM.LOCAL.CAT.VAL [ SUBJ < [ LOCAL [ CONT.HOOK #hook,
                                                CAT [ NCORP-MOD #nc-mod,
                                                    VAL #val,
                                                    HEAD +np ] ] ] >,
                              COMPS #comps ],
      DTR.SYNSEM.LOCAL.CAT.VAL [ SUBJ < [ LOCAL [ CONT.HOOK #hook,
                                                CAT [ NCORP-MOD #nc-mod,
                                                    VAL #val ] ] ] ] >,
                              COMPS.REST #comps ] ].
```

This rule is underspecified when it comes to the subject’s case value. I then check if the user defined a new case value for the subject. If it has changed, I access this value and assign it to the case value on the mother’s subject with the TDL line in (73) where `a_case` is the a choice made by the user and found in the `red-val-intrans-yes-case` variable in the choices file. If not, one of two things happens. If case is even present in the language, then the subject’s case value is identified between mother and daughter with the TDL in (74). If not, nothing is added to this rule.

---

<sup>9</sup>When fully realized in the Tongan grammar, the DTR feature of this lexical rule is constrained to be of type *NI-valence-rule-dtr*, which allows the transitive verb in (69) to unify with this feature.

(73) := [ SYNSEM.LOCAL.CAT.VAL.SUBJ < [ LOCAL.CAT.HEAD.CASE ' + a\_case + ' ] > ].

(74) := [ SYNSEM.LOCAL.CAT.VAL.SUBJ < [ LOCAL.CAT.HEAD.CASE #case ] >,  
DTR.SYNSEM.LOCAL.CAT.VAL.SUBJ < [ LOCAL.CAT.HEAD.CASE #case ] > ].

Based on the interaction with the valence change library (Curtis, 2018), I altered my analysis for the `reduce-lex-rule` so that instead of just defining an empty COMPS list on the mother, the rule instead identifies COMPS.REST on the daughter with COMPS on the mother. Therefore, if there is only one element on the daughter's COMPS list, the result is the same, but if there was an added element, for example a demoted subject at the end of the daughters COMPS list, it will be identified with the sole element on the mother's COMPS list, and the first element on the daughter's will not be identified with anything.

If the user chose the transitive reduction behavior option, then the following TDL lines are added to `mylang.tdl` to include this lexical rule type. This TDL corresponds to the lexical rule given in (61) in the previous chapter.

(75) `reduce-lex-rule` := `no-ccont-lex-rule` &  
[ SYNSEM.LOCAL.CAT.VAL [ SUBJ #subj,  
COMPS #comps &  
[ FIRST.OPT + ] ],  
DTR.SYNSEM.LOCAL.CAT.VAL [ SUBJ #subj,  
COMPS #comps ] ].

Additionally, if transitive reduction NI is indicated to exist in the language, the TDL in (76) is added to the matrix type addenda at the top of `mylang.tdl`.

(76) `basic-head-1st-comp-phrase` :+  
[ HEAD-DTR.SYNSEM.LOCAL.CAT.VAL.COMPS < [ OPT - ] > ].

This addition states that the head daughter of a `head-comp` rule must have an element on its COMPS list that is assigned to be [OPT -]. In other words, there must be an overt complement in the syntax for this phrase structure rule to apply. This ensures that the output of the lexical rule in (75) does not feed a `head-comp` rule, and rather that the non-branching phrase structure rule for optional complements will apply.

Regardless of the which lexical rule type with the name `reduce-lex-rule` is added to `mylang.tdl`, it is then instantiated in `lrules.tdl` since it is non-inflecting.

### *Promote lexical rules*

For both of the Promotion NI lexical rules defined in my analysis, implementation mostly involves simply adding the lexical rules from my analysis in TDL format as lexical rule types. The one additional action needed is to add the PRED value of the relation linking the promoted oblique argument to the verb's event. If the user selected the check box stating that an oblique argument can be promoted to direct object position, the following TDL is added as a lexical rule type to `mylang.tdl` which corresponds to my analysis of promoted obliques in (64) in the previous chapter.

```
(77) promote-obl-lex-rule :=
      [ SYNSEM.LOCAL.CAT.VAL [ SUBJ #subj,
                              COMPS [ FIRST [ OPT -,
                                             LOCAL [ CAT [ NCORP-MOD -,
                                                         VAL.SPR < >,
                                                         HEAD noun &
                                                         #head ],
                                                         CONT.HOOK.INDEX #arg2 ] ],
                              REST #rest ] ],
      DTR.SYNSEM.LOCAL [ CAT.VAL [ COMPS [ FIRST.LOCAL.CAT.HEAD #head,
                                           REST #rest ],
                                     SUBJ #subj ],
                        CONT.HOOK.INDEX #arg1 ],
      C-CONT [ RELS.LIST < arg12-ev-relation &
               [ ARG1 #arg1,
                 ARG2 #arg2 ] >,
              ICONS.LIST < >,
              HCONS.LIST < > ] ].
```

Then, based on the value stored in the “ni-predname” variable in the choices file, the following line of TDL is inserted into the lexical rule type, customized with the user's input.

```
(78) promote-obl-lex-rule :=
      [ C-CONT.RELS.LIST < [ PRED ‘‘ + ni-predname + \_rel’’ ] > ]
```

The customization system merges statements about the same type, so in the output grammar, both of these TDL snippets will appear in the same lexical rule type. For instance, if the user had input ‘benefactive’ as the predicate value in the customization system, the relation added to C-CONT.RELS would look like the following.

```
(79) C-CONT.RELS.LIST < arg12-ev-relation &
      [ ARG1 #arg1,
        ARG2 #arg2,
        PRED "benefactive_rel" ] >
```

If the user indicated that it is also possible for the possessor of the incorporated noun to appear as the direct object of the verb, the following lexical rule type will be added to `mylang.tdl`. This lexical rule type corresponds to my lexical rule for possessor raising in (63) in the previous chapter.

```
(80) promote-poss-lex-rule :=
      [ SYNSEM.LOCAL.CAT.VAL [ SUBJ #subj,
                              COMPS [ FIRST [ OPT -,
                                              LOCAL [ CAT [ NCORP-MOD #nc-mod,
                                                            VAL.SPR < >,
                                                            HEAD noun &
                                                            #head ],
                                                            CONT.HOOK.INDEX #arg2 ] ],
                              REST #rest ] ],
      DTR.SYNSEM.LOCAL.CAT.VAL [ SUBJ #subj,
                              COMPS [ FIRST [ LOCAL [ CAT [ NCORP-MOD #nc-mod,
                                                            HEAD #head ],
                                                            CONT.HOOK [ INDEX #arg1,
                                                            LTOP #lbl ] ] ],
                              REST #rest ] ],
      C-CONT [ RELS.LIST < arg12-ev-relation &
              [ PRED "poss_rel",
                LBL #lbl,
                ARG1 #arg1,
                ARG2 #arg2 ] >,
              ICONS.LIST < >,
              HCONS.LIST < > ] ].
```

As with other translations of analysis into implemented TDL, I have accounted for COMPS lists of variable length in the TDL above by referencing the first element on the

COMPS list, and copying up the rest.

### *Doubling and Stranding*

In my analysis, I defined one unified lexical rule for verbs with an incorporated noun to be able to combine syntactically with some modifying element outside of the verb whether that be a more specific noun, an adjective, or a demonstrative. In implementation, I have compartmentalized Doubling and Stranding as two different operations in order to account differing co-occurrence restrictions between the two types, and also to allow for more specific feature constraints. The result of this separation is that users may define forbid constraints for one type and not the other in the Customization System. I am also able to account for the fact that in languages like Tongan [ton] in which Reduction NI constructions behave intransitively, stranding constructions appear to as well.<sup>10</sup>

If the user indicated that doubling occurs in the language, the following TDL is added to `mylang.tdl`. This corresponds to my lexical rule for doubling in (65) in the previous chapter.

```
(81) double-noun-lex-rule :=
      [ SYNSEM.LOCAL.CAT.VAL [ SUBJ #subj,
                              COMPS [ FIRST [ OPT -,
                                          LOCAL [ CAT [ NCORP-MOD +,
                                                    HEAD noun ],
                                                    CONT.HOOK [ LTOP #ltop,
                                                                INDEX #ind ] ] ],
                                          REST #rest ] ],
      DTR.SYNSEM.LOCAL.CAT.VAL [ SUBJ #subj,
                              COMPS [ FIRST.LOCAL.CONT.HOOK [ LTOP #ltop,
                                                                INDEX #ind ],
                              REST #rest ] ],
      C-CONT [ RELS.LIST < >,
              HCONS.LIST < >,
              ICONS.LIST < > ] ].
```

In this lexical rule type, the HEAD feature on the COMPS list is constrained to have a

---

<sup>10</sup>For more information, see (6.1.5) in the evaluation chapter.

value of *noun*, making it appropriate for verbs which will combine with a phrase headed by a noun rather than an adnominal modifier.

The same choice also triggers the addition of the following phrase structure rule to the ‘phrases’ section of `mylang.tdl`. This corresponds to the unary phrase structure rule in (66) from the previous chapter which allows doubled nouns to modify the IN.

```
(82) type-ni-mod-phrase := unary-nonloc-phrase & head-only &
      [ SYNSEM.LOCAL.CAT [ WH #wh,
                          VAL #val,
                          NCORP-MOD + ],
        HEAD-DTR.SYNSEM.LOCAL [ CAT [ HEAD noun,
                                      WH #wh,
                                      VAL #val &
                                      [ SPR < > ],
                                      NCORP-MOD - ],
                                  CONT.HOOK.INDEX #arg2 ],
        C-CONT [ RELS.LIST < arg12-ev-relation &
                  [ PRED "type_rel",
                    LBL #lbl,
                    ARG1 #ind,
                    ARG2 #arg2 ] >,
                  ICONS.LIST < >,
                  HCONS.LIST < >,
                  HOOK [ LTOP #lbl,
                        INDEX #ind ] ] ].
```

`double-noun-lex-rule` is instantiated in `lrules.tdl`, and `type-ni-mod-rule` is instantiated in `rules.tdl` because they are a non-inflecting lexical rule and a phrase structure rule, respectively.

The same process happens if the user indicates that stranded modifiers of an IN appear in the language. TDL for the lexical rule type and the phrasal type in (83) and (84) below are added to `mylang.tdl`, and instantiated where appropriate in `lrules.tdl` and `rules.tdl`. These correspond to (67) and (68) in the previous chapter — the lexical and phrase structure rule pair which allows stranded modifiers to combine with an incorporating verb and modify the IN.

```
(83) strand-mod-lex-rule :=
```

```

[ SYNSEM.LOCAL.CAT.VAL [ SUBJ < [ LOCAL [ CONT.HOOK #hook,
                                     CAT [ NCORP-MOD #nc-mod,
                                     VAL #val,
                                     HEAD +np ] ] ] >,
  COMPS [ FIRST [ OPT -,
                LOCAL [ CAT [ NCORP-MOD +,
                            HEAD +jp ],
                CONT.HOOK [ LTOP #ltop,
                            INDEX #ind ] ] ] ],
  REST #rest ] ],
DTR.SYNSEM.LOCAL.CAT.VAL [ SUBJ < [ LOCAL [ CONT.HOOK #hook,
                                     CAT [ NCORP-MOD #nc-mod,
                                     VAL #val ] ] ] >,
  COMPS [ FIRST.LOCAL.CONT.HOOK [ LTOP #ltop,
                                INDEX #ind ],
  REST #rest ] ],
C-CONT [ RELS.LIST < >,
        HCONS.LIST < >,
        ICONS.LIST < > ] ].

```

- (84) `adj-ni-mod-phrase := unary-nonloc-phrase & head-only &`  

```

[ SYNSEM.LOCAL.CAT [ WH #wh,
  VAL [ SUBJ #subj,
        COMPS #comps,
        SPEC #spec,
        SPR < > ],
  NCORP-MOD + ],
HEAD-DTR.SYNSEM.LOCAL [ CAT [ HEAD +jp &
  [ MOD.FIRST.LOCAL.CONT.HOOK.INDEX #arg1 ],
  WH #wh,
  VAL [ SUBJ #subj,
        COMPS #comps,
        SPEC #spec ],
  NCORP-MOD - ],
  CONT.HOOK.LTOP #ltop ],
C-CONT [ RELS.LIST < >,
        ICONS.LIST < >,
        HCONS.LIST < >,
        HOOK [ LTOP #ltop,
              INDEX #arg1 ] ] ].

```

In (83), the HEAD value of the head daughter is constrained to be *+jp*. This is a disjunctive type for adjectives and adpositions found in `head-types.tdl`. This specification

accounts for different types of modifiers such as stranded modifiers with different HEAD values such as adpositional phrases.<sup>11</sup>

Additionally, if the user indicated that Reduction NI results in a case change on the subject, the same TDL lines are added to the lexical rule type for stranded modifiers. If no case change is relevant, the case value is identified between mother and daughter.

If either doubling or stranding appears in the language, the NCORP-MOD feature needs to be constrained to only be positive on those phrases which have undergone either the `type-ni-mod-phrase` or the `adj-ni-mod-phrase` rules. To achieve this, I define the NCORP-MOD values to be negative everywhere else in the grammar, as follows:

In the matrix type addenda, the definition for `lex-item` receives the following changes.

```
(85) lex-item := [ SYNSEM.LOCAL.CAT [ NCORP-MOD -,
                                VAL [ SUBJ nc-list,
                                       COMPS nc-list,
                                       SPR nc-list,
                                       SPEC nc-list ] ] ] .
```

Any element which inherits from `lex-item` is constrained to be `[NCORP-MOD -]`. Additionally, all of its valence lists are of type *nc-list*, which is a new parameterized list I defined in `matrix.tdl`. Parameterized lists are definitions which allow a list to either be empty or to have one or more elements, and to define constraints on the elements in the case that there are any. For the *nc-list*, if there are elements on the list, they are thus constrained to have a negative NCORP-MOD value by the following definition.

```
(86) nc-cons := nc-list & cons &
        [ FIRST synsem & [ LOCAL.CAT.NCORP-MOD - ],
          REST nc-list] .
```

I also make the following addition to the definition for the `bare-np` rule to make sure that one of the NCOPR-MOD phrase structure rules cannot feed the `bare-np` rule.

---

<sup>11</sup>The decision to include both adjectives and adpositions comes from stranded prepositional phrases in Tongan. For more information, see (6.1.5) in the following chapter.

(87) `bare-np-phrase :=`  
`[ SYNSEM.LOCAL.CAT.NCORP-MOD -`  
`HEAD-DTR.SYNSEM.LOCAL.CAT.NCORP-MOD - ] .`

Finally, I make additions to the phrase structure rules independently present in the language’s grammar to specify negative values for the NCORP-MOD feature on the head daughters.<sup>12</sup> This includes any variation of the head-comp rule as well as any `head-adj-int` rules if applicable. All of these updates ensure that the only time the NCORP-MOD value is positive on some element is due to one of the non-branching rules defined in (82) and (84) above, and that the value is not left underspecified elsewhere in the grammar.

### 5.3.3 Summary

In this section, I described how I translated the analyses of NI laid out in the previous chapter into TDL, and how I ensure that the correct TDL is included in a user’s grammar based on their answers to the elicitation questions outlined in the previous section. I showed how I was able to automate some of the work for users with a single position class approach. Since my analysis of NI involves a pair of lexical rules, I designed my system to allow the user to define a single PC with IS-LRTs and IS-LRIs, and create the second ‘ghost’ PC behind the scenes. In the output grammars, the second PC has non-inflecting instances of lexical rule types for each type of NI indicated to be present in the language. I also described each of the additions to other types in the grammar and to `matrix.tdl` needed to constrain relevant features and make the system work.

## 5.4 Validation

In implementing my library, I also designed a series of validation tests to interactively guide users to make choices on the customization system that will produce valid grammars, and to

---

<sup>12</sup>Specifying this negative value on the head daughters of phrase structure rules was sufficient in constraining this feature during development, however future work should further ensure that the value is negative on the mother of these rules by identifying the value between mother and head daughter, for example.

block them from making choices which will cause an error as the Grammar Matrix creates a customized grammar for the user. Validation also warns users if the choices made are theorized to not be a combination of behaviors found in the world's languages. My validation tests for the most part remind the user to fill out all of the necessary fields on the customization subpages. In this section, I will describe the errors and warnings present on the NI page.

For the most part, errors on the NI subpage can be thought of as 'right-leaning' meaning that they refer to the nested questions on the subpage. For example, if noun incorporation is indicated to occur in the language (i.e. the user has selected the first checkbox on the top of the NI subpage), then they must select one or more of the NI types (Reduction, Promotion, etc.). If the user only selects the first checkbox and then clicks the 'save and stay' button, a red asterisk will appear on the page which, when hovered over, tells the user that they must chose at least one of the three types of NI. Similarly, each type of NI elicits more information from the user in order to determine which lexical rule types to add to the grammar. If these are left unanswered, grammar compilation is not possible, so an error message is displayed which must be addressed before the user is able to download their grammar. An example is shown in Figure 5.8.

Optional questions are not required to be answered. For example, the user may define a forbid constraint for one or more types of NI, but leaving this field blank is also valid and will not produce an error message. I differentiate between optional and required fields in prose on the subpage with the prompt 'If so, ...' to indicate that the user should continue answering questions. However, even if they miss this, the validation error will ensure it is answered before moving on.

There are two spots on the page where the user is prompted to input contentful answers rather than just selecting a check box or radio button: the new subject case value for intransitive Reduction NI, and the predicate value for promoted oblique arguments. If either of these inputs are relevant to the grammar (i.e. reduction NI occurs, and exhibits intransitive behavior; or promotion NI occurs, and the promoted argument is an oblique argument), a

## Noun Incorporation [\[documentation\]](#)

Noun Incorporation is the process of morphologically incorporating a noun, usually the direct object of the verb, into the verb itself resulting in a complex verb.

Noun Incorporation occurs in this language.

Do direct objects of transitive verbs incorporate? If so, what happens to the argument structure of the verb? Check all strategies that apply. A second position class for noun incorporation will be created behind the scenes with lexical rule types for each NI strategy that you select below. This position class and its associated lexical rule types will not be visible on the customization page. You might want to define a forbid constraint for one NI strategy but not another as you would do for LRTs or PCs on the morphotactics page. You will have the option to select previously defined PCs or LRTs to forbid with each selected NI type below.

The direct object of the verb incorporates and there is no other overt expression of this argument.

If so, how does the verb behave?

\*  The verb appears transitive (i.e. transitive morphology, case)  
 The verb appears intransitive (i.e. intransitive morphology, case)

You must select whether reduction type NI constructions behave transitively or intransitively.

checked on the subject change?

No

Yes. New case:  ▼

Are there any position classes that this type of NI should not co-occur with?

Figure 5.8: A validation error showing that the user has not finished describing the behavior of Reduction NI in the language

validation error will occur if the user does not provide the information requested by these two questions.

Validation warnings, on the other hand, can be thought of as ‘left-leaning’ in that they will appear if a user has answered a nested question, but did not chose the outer checkbox prompting this nested question. In this case, making this choice will not keep the grammar from compiling, but the choice itself will be ignored by the code processing the choices file. Since the entire subpage is visible to the user all at once, it is possible to make any choice at any time. For example, it is possible to check the box for promoted possessors without checking the box that Promotion NI occurs at all. If this choice is made, due to how the backend code evaluate the choices file, this choice will never even be considered. The resulting grammar would not contain the promote-poss-lex-rule, nor any promotion rules at all. The warning serves to remind the user to make the outer choice if that is what they intended to do. Similarly, nothing related to NI will happen if the first check box on the page is not selected. An example of a validation warning is shown in Figure 5.9.

Finally, development with the Mohawk [moh] grammar led me to create an additional validation warning concerning the co-occurrence of some incompatible position classes.<sup>13</sup> In Mohawk, intransitive verbs are input to an intransitive agreement affix, and transitive verbs input to a transitive agreement affix. Incorporated nouns appear between the verb root and agreement morphology and only take in transitive verbs as input. Since some types of NI behave intransitively (i.e. Reduction NI) and other types transitively (i.e. Promotion NI), this PC which houses all of the NI lexical rule types was defined to be input to both agreement affix PCs. Due to the way the Grammar Matrix handles input PCs and LRTs, the grammar was producing parses where intransitive verb roots were feeding the transitive agreement lexical rules. This is a limitation of my library, but is easily fixed via grammar engineering and hand-editing TDL. In the documentation for my library, I include some ideas of how to handle cases like this if it occurs in the user’s project. I also include a validation

---

<sup>13</sup>The Mohawk grammar is described in more detail in evaluation in chapter 6.

## Noun Incorporation [\[documentation\]](#)

Noun Incorporation is the process of morphologically incorporating a noun, usually the direct object of the verb, into the verb itself resulting in a complex verb.

Noun Incorporation occurs in this language.

Do direct objects of transitive verbs incorporate? If so, what happens to the argument structure of the verb? Check all strategies that apply. A second position class for noun incorporation will be created behind the scenes with lexical rule types for each NI strategy that you select below. This position class and its associated lexical rule types will not be visible on the customization page. You might want to define a forbid constraint for one NI strategy but not another as you would do for LRTs or PCs on the morphotactics page. You will have the option to select previously defined PCs or LRTs to forbid with each selected NI type below.

?  The direct object of the verb incorporates and there is no other overt expression of this argument.

You have made a choice about Reduction type NI. Please state that it occurs in the language.

- The verb appears transitive (i.e. transitive morphology, case)  
 The verb appears intransitive (i.e. intransitive morphology, case)

If intransitive, does the case marked on the subject change?

- No  
 Yes. New case:  ▼

Are there any position classes that this type of NI should not co-occur with?

Figure 5.9: A validation warning showing that a nested choice was made without choosing the corresponding checkbox

warning to alert the user that the PC for NI is taking inputs that are not compatible with all of the PCs that it itself is feeding, which may or may not cause an issue elsewhere in the grammar.

## **5.5 Summary**

In this chapter, I described how I designed my user-facing web questionnaire to elicit the range of behavior of NI in a given language. I then demonstrated how I use the information gathered from the customization system via the choices file to determine which parts of my NI analysis are relevant to the language in question. I showed how the feature structures from chapter 4 map to TDL found in the various grammar files, and how interaction with other libraries fine-tuned my analysis during implementation. Finally, I described how I use various validation tests to warn users about potentially unintended outcomes of the choices made, and ensure that by the time they download their grammar, that each choice regarding NI will be compatible with the customization program. In the following chapter, I will discuss how I evaluated this system during and after development with natural and psuedo-languages which exhibit different types of NI.

## Chapter 6

# EVALUATION

This chapter describes the process of evaluating my analysis of Noun Incorporation and its implementation into the Grammar Matrix.<sup>1</sup> I use three different classes of test languages to evaluate the performance of my system: illustrative languages, pseudo-languages, and held-out languages. As I developed my analysis and implementation, I created grammars for five real languages with noun incorporation using a non-published version of the customization system which includes my additions and changes. I then constructed a suite of pseudo-languages to test the combinations of choices not represented by the illustrative languages. At this point, I froze the development of the system and tested it against five unseen real languages which I did not consider when developing my analysis.

For each language, I collected a testsuite of strings representing grammatical and ungrammatical sentences in the language based on the literature. For the illustrative and held out languages, these sentences come from a combination of examples pulled directly from literature, simplified examples from the literature, and sentences I constructed based on my understanding of the facts of the language. Any changes to the sentences in my testsuites from what appeared in the literature are described in the following sections, but are only done when the functionality to model the sentence as is is not supported by other elements of the customization system. For the pseudo-languages, I created the testsuites from scratch using readable but simplified notation.

The metrics for which I evaluate my system are coverage, overgeneration, and ambiguity. Coverage refers to the ability to correctly parse grammatical sentences while overgeneration

---

<sup>1</sup>The version of the system that was used for this evaluation can be found at: <https://github.com/delphin/matrix/commit/c01672235357b2e4aa97795046f341816bc0335b>.

is the unwanted parsing of ungrammatical sentences. A successful system should maximize coverage and minimize overgeneration. Ambiguity refers to the number of parses the grammar licenses for a given string. Sometimes, a grammar may provide multiple different valid parses for the same string, but in general, I strive to reduce spurious ambiguity and limit the number of generated parses per sentence.

Section 6.1 presents my work with five illustrative languages, section 6.2 describes how I rounded out the coverage of development tests with seven pseudo-languages, and section 6.3 shows the performance of my system on five previously unconsidered, held-out languages.

### **6.1 *Illustrative Languages***

This section describes the five natural languages displaying Noun Incorporation that I considered when developing my analysis of NI and working towards a concise implementation. These languages are Chukchi [ckt] (Chukotko-Kamchatkan), Southern Tiwa [tix] (Kiowa-Tanoan), Mapudungun [arn] (Araucanian), Mohawk [moh] (Iroquoian), and Tongan [ton] (Austronesian). I chose these languages as they were featured prominently in the NI literature and illustrate the diversity of the range of expressing NI cross-linguistically. The distribution of the types of NI present in each language is shown in Table 6.1. The columns in this table are grouped by type of NI: Reduction, Promotion, and Modifier. Reduction NI is further divided into three columns: trans, intrans same, and intrans different. ‘trans’ and ‘intrans’ refer to whether the language behaves transitively or intransitively as a result of Reduction NI. ‘same’ and ‘different’ refer to whether the case value marked on the subject changes as a result of intransitive Reduction NI. ‘promote poss’ and ‘promote obl’ refer to promoted possessives and oblique arguments, respectively. Additional facts about the languages are shown in Table 6.2, including information about their word orders, case systems, verbal agreement affixes, (absence of) transitivity markers, and valence changing operations. In this section, I introduce each language and the grammar I created to test and refine my analysis. For each language, I report the performance of the grammar over the constructed testsuite. When evaluating the parses licensed by the grammar, I hand inspect each tree

structure and semantic representation to ensure that they are the expected outputs.

Language	Reduction NI			Promotion NI		Modifier NI	
	trans	intrans same	intrans different	promote poss	promote obl	strand	double
Chukchi			✓	✓	✓		
S. Tiwa	✓					✓	
Mapudungun		✓		✓			
Mohawk		✓			✓	✓	✓
Tongan			✓			✓	

Table 6.1: Illustrative language choices on NI subpage

Language	WO	case system	subj/obj agreement	transitivity marking	valchange ops
Chukchi	free	erg/abs, locative	subj+obj	no	applicative
S. Tiwa	free	(nom/acc), locative	subj+obj	no	passive
Mapudungun	SVO	none	subj, obj	no	applicative
Mohawk	free	none	subj+obj	no	
Tongan	free	erg/abs	none	no	

Table 6.2: Illustrative language additional choices for testing library interaction

### 6.1.1 *Chukchi [ckt]*

Chukchi (also spelled Chukchee in some literature) is a Chukotko-Kamchatkan language spoken in eastern Russia. The language displays ergative absolutive alignment with these two main cases marked morphologically on overt core arguments. Verbs are inflected with a

single agreement affix which agrees in person and number with the subject and object. To test the Chukchi grammar, I created a testsuite with seven grammatical and five ungrammatical sentences. The grammatical sentences come from Gerdts (2017) and Baker et al. (2005) while I personally created the ungrammatical ones. Some are described below.

Chukchi has Reduction NI where the resulting sentence is intransitive. The subject in Reduction NI constructions is inflected with absolutive case marking, and the agreement marker is from the intransitive paradigm, agreeing only with the subject. This pattern is shown in the alternation of the transitive sentence in (88a) and the NI construction in (88b).<sup>2,3</sup>

- (88) a. *etlege qorane temnen*  
 etleg-e qorane-en tem-nen  
 father-ERG reindeer-ABS kill-3SGA.3SGO  
 ‘The father killed the/a reindeer.’ [ckt](Polinsky, 1990 via Gerdts, 2017:88)
- b. *etlegen qaanmege*  
 etleg-en qaa-nme-ge  
 father-ABS reindeer-kill-3SS  
 ‘The father killed the reindeer.’ [ckt](Polinsky, 1990 via Gerdts, 2017:88)

Chukchi also allows the promotion of both possessors (89) and oblique arguments (90) with NI. Both promoted arguments are marked with the absolutive case while the subject is marked with ergative case. The agreement affix agrees with both the subject and the promoted object.<sup>4</sup>

---

<sup>2</sup>I added the absolutive affix on the noun *qorajə* for simplicity. All other absolutive nouns in the source text were inflected with this absolutive morpheme and it was not explained why this noun was not. The original gloss does indicate that this noun is absolutive.

<sup>3</sup>In this chapter, I represent the sentences in a simplified orthography for ease of use with the LKB software. For example, these sentences were presented in the Analysis chapter in IPA but are translated into a simplified Latin script.

<sup>4</sup>Both the IN and the promoted object are 3rd person singular nouns, but the fact that the agreement marker agrees only with the subject in Reduction NI constructions like (88b) suggests that the agreement here is with the promoted noun, not the IN.

- (89) *enanen pojgemcatkonen remkele*  
 enan-e pojge-mcatko-nen remkel-en  
 they-ERG spear-break-3SGS.3SGO guest-ABS  
 ‘They broke the guest’s spear.’ [ckt] (Spencer, 1995:450 via Baker et al., 2005:171)
- (90) *tumge ekek kaynmenen*  
 tumg-e ekek-en kayn-nme-nen  
 friend-ERG son-ABS bear-kill-3SGS.3SGO  
 ‘The friend killed the bear for his son.’ [ckt] (Polinsky, 1990 via Gerdts, 2017:91)

Chukchi has ditransitive verbs where the second object is marked with a locative case marker as in (91). While the Grammar Matrix customization system does not yet support ditransitive verbs, I was able to model this feature of the language with a non-inflecting applicative lexical rule which constrains the case to be locative on the added object.<sup>5</sup>

- (91) *etlege metqemet kawkawek kilinin*  
 etleg-e metqe-en kawkaw-ek kili-nen  
 father-ERG butter-ABS bread-LOC spread-3SGS.3SGO  
 ‘The father spread the butter on the bread.’ [ckt] (Polinskaja and Nedjalkov, 1987:240 via Baker et al., 2005:153)

The absolutive argument in a ditransitive sentence can be incorporated. When this happens, the subject and second object can either be marked with the intransitive absolutive and locative cases as in (92a) or with the transitive ergative and absolutive case as in (92b).

- (92) a. *etlegen kawkawek metqerkelege*  
 etleg-en kawkaw-ek metqe-rkele-ge  
 father-ABS bread-LOC butter-spread-3sS  
 ‘The father spread the butter on the bread.’ [ckt] (Polinskaja and Nedjalkov, 1987:240 via Baker et al., 2005:153)
- b. *etlege kawkaw metqerkelenen*  
 etleg-e kawkaw-en metqe-rkele-nen  
 father-ERG bread-ABS butter-spread.on-3SGS.3SGO

---

<sup>5</sup>This modeling choice is not meant to be an analysis of ditransitives in Chukchi. Rather, it is an expedient in order to test the analysis of NI as presented.

‘the father spread the butter on the bread.’ [ckt] (Polinskaja and Nedjalkov, 1987 via Gerdt, 2017:91)

I account for this alternation with the use of two different inflection strategies. I analyze the sentence in (92a) as a combination of the applicative rule adding a second object with locative case and the **reduce-lex-rule** incorporating the IN. This is consistent with the absolutive marking on subjects in other Reduction constructions. (92b), on the other hand, I analyze as a result of the oblique argument promotion rule. The ergative/absolutive case marking in this sentence are likewise consistent with other Promotion constructions.

The results for coverage, overgeneration, and parsing ambiguity for the Chukchi grammar with the given testsuite are shown in Table 6.3. 7/7 grammatical sentences correctly parsed and 5/5 ungrammatical sentences did not. There was an average of 1.43 parses per sentence, due to the presence of both types of Promotion NI. The two rules structurally work the same way, but differ in the semantic representations they produce. Thus, for languages which allow both types of Promotion NI, the resulting ambiguity is valid.

### 6.1.2 *Southern Tiwa [tix]*

Southern Tiwa is a Kiowa-Tanoan language spoken in New Mexico. I created a testsuite to evaluate the performance of the Southern Tiwa grammar consisting of 13 grammatical and 6 ungrammatical sentences. The grammatical sentences come from Allen et al., 1984. Ungrammatical sentences come from Baker Baker, 1988, Allen et al., 1984, and myself.

Southern Tiwa is the only one of the illustrative languages to display transitive features for Reduction NI evidenced by agreement morphology agreeing with the incorporated noun. A single morpheme agrees in person and number with the subject and person, number, and animacy with the object. Examples (93a) and (93b) show how a change in object agreement inflection changes the meaning of the sentence.

- (93) a. *tiseuanmuban*  
 ti-seuan-mu-ban  
 1SGS.3SGO-man-see-past

‘I saw the/a man.’ [tix](Allen et al., 1984:294)

- b. *biseuanmuban*  
 bi-seuan-mu-ban  
 1SGS.3PLO-man-see-past  
 ‘I saw men.’ [tix] (Allen et al., 1984:295)

Southern Tiwa also allows external modifiers to modify an incorporated noun. For example, the demonstrative in (94) modifies the IN, and also agrees with it in number.

- (94) *yede adirukarhi*  
 yed-e a-diru-kar-hi  
 that-SG 2SGS.AO-chicken-eat-FUT  
 ‘You will eat that chicken.’ [tix] (Allen et al., 1984:296)

In Southern Tiwa, inanimate nouns must be incorporated when they are direct objects as shown in the ungrammatical example (95). In order to account for mandatory incorporation, I created a class of inanimate nouns defined to have nominative case. I specified that the all other nouns have accusative case. Since case marking is not realized morphologically or adpositionally in Southern Tiwa, these specifications did not negatively impact the rest of the grammar. I then constrained transitive verbs to only take complements with accusative case. Therefore, inanimate nouns could never combine with transitive verbs as an object via the *head-comp* rule. The rationale for using the *CASE* feature for this purpose is that a *CASE* system constrains which nouns can occur in which argument positions.

- (95) *\*shut tipeban*  
 shut ti-pe-ban  
 shirt 1SGS.AO-make-PAST  
 ‘I made the shirt.’ [tix] (Allen et al., 1984:293)

There is a passivizing morpheme *-e* in Southern Tiwa which demotes the subject to object marked with the instrumental case and promotes the object to subject as in (96a). The resulting object can also be incorporated while maintaining a semantic paraphrase with the unincorporated passive construction as in (96b).

- (96) a. *khwienide edeureban kanideba*  
 khwien-ide edeuri-e-ban kan-ide-ba  
 dog-SG kick-PASS-PAST horse-SG-inst  
 ‘The dog was kicked by the horse.’ [tix] (Allen et al., 1984:302)
- b. *khwienide kanedeuriaban*  
 khwien-ide kan-edeuri-e-ban  
 dog-SG horse-kick-PASS-PAST  
 ‘The dog was kicked by the horse.’ [tix] (Allen et al., 1984:302)

When defining position classes on the morphology subpage of the customization system, I had to make sure that the passive rule was input to the noun incorporation rule and not the other way around in order to incorporate the correct element. In the end, my system is able to correctly parse the sentence in (96b) which constitutes a successful interaction between my library and the valence change library (Curtis, 2018).

In addition to stranded demonstratives, INs in Southern Tiwa can also strand the second half of a coordinated noun phrase seen in (97). I do not account for the stranding of elements other than adjectives and adpositional phrases,<sup>6</sup> so this sentence does not parse and accounts for the less than perfect performance in terms of coverage for this language’s grammar.

- (97) *tikantuwiban ba bakaidean*  
 ti-kan-tuwi-ban ba khwien-ide-an  
 1SGS.AO-horse-buy-PAST and dog-SG-with  
 ‘I bought a horse and a cow.’ [tix] (Allen et al., 1984:301)

The results for the Southern Tiwa grammar are shown in Table 6.3. As mentioned, 1/13 grammatical sentences did not parse due to the stranded coordinated NPs being out of scope for this thesis. 6/6 ungrammatical sentences failed to parse. The two grammatical sentences in the testsuite involving passivization each received four parses. This ambiguity stems from a bug with the valence change library under-constraining the valence lists of added arguments. Thus, the ambiguity is not a result of the noun incorporation library.

---

<sup>6</sup>This coordination strategy uses the coordinating word *ba*. I analyze Southern Tiwa as an ‘AND-language’ in terms of coordination (Drellishak and Bender, 2005). For an example of a ‘WITH-language’ coordination strategy which is licensed by this library, see the discussion on Tongan [ton] in section (6.1.5) below.

### 6.1.3 Mapudungun [arn]

Mapudungun is an Araucanian language spoken by the Mapuche people in Chile and Argentina. This language has an SVO word order and no case marking for core or oblique arguments. In transitive sentences, there are separate agreement affixes which agree in person and number with the subject and the object. I created a testsuite with seven grammatical and four ungrammatical sentences. Each of these sentences—grammatical and ungrammatical—comes from Baker et al., 2005.

When Reduction NI occurs as in (98), only the subject agreement affix appears indicating intransitive behavior. The lack of a case marking system by default entails that the case on the subject does not change due to Reduction NI.

- (98) *Ni chao kintuwakaley*  
 Ni chao kintu-waka-le-y  
 my father seek-cow-PROG-IND.3SG.S  
 ‘My father is looking for cows.’ [arn] (Salas, 1992:195 via Baker et al., 2005:139)

Incorporation in Mapudungun can also allow the promotion of the possessor of the IN to object position as in (99). Promotion results in a transitive sentence with both subject and object agreement on the verb, with the object agreement agreeing with the promoted possessor.<sup>7</sup>

- (99) *Juan ngillawakafiy Pedro*  
 Juan ngilla-waka-fi-y Pedro  
 Juan buy-cow-3O-ind.3SG.S Pedro  
 ‘Juan bought Pedro’s cow.’ [arn] (Baker et al., 2005:167)

Applicative morphology, namely the affix *-lel*, adds an object to the verb’s COMPS list. When used in conjunction with NI, the resulting sentence is transitive, again with subject and object agreement morphology as seen in (100).

---

<sup>7</sup>The object agreement morpheme in fact only specifies third person which is consistent with both the IN and the raised possessor. The absence of an object agreement marker in sentences like (98) supports the analysis of this morpheme agreeing with the promoted object.

- (100) *Juan ngillawakalelfy*  
 Juan ngilla-waka-lel-fi-y  
 Juan buy-cow-APPL-3O-IND.3SG.S  
 ‘Juan bought a cow for him.’ [arn] (Baker et al., 2005:140)

This sentence can be realized with the **reduce-lex-rule** removing the first element from the COMPS list, and the applicative rule adding one after for a resulting COMPS list with one element. The order of affixes on the verb makes this interaction possible with the IN appearing closer to the verb root than the applicative morpheme.

The results of the Mapudungun grammar on the created testsuite are shown in Table 6.3. All of the grammatical sentences parsed with no ambiguity. The one ungrammatical sentence which parsed does have an alternative reading. The ungrammatical string was constructed to show that doubling does not occur in the language. The **promote-lex-rule** is used instead to create a valid representation of the string. Thus, the grammar does work as intended.

#### 6.1.4 Mohawk [moh]

Mohawk is a member of the Iroquoian language family and is spoken in the Canadian states of Ontario and Quebec as well as New York state in the U.S.. There is no dominant word order in Mohawk nor case marking on core arguments. A single verbal affix agrees in person, number, and gender with both the subject and the object in transitive sentences, and just the subject in intransitive sentences. The testsuite for Mohawk has 15 grammatical and six ungrammatical sentences. The grammatical sentences come from Baker et al., 2005 and Baker, 1996. I created the ungrammatical sentences by hand.

Like Mapudungun, Reduction NI in Mohawk results in an intransitive sentence with no case change on the subject since no case system exists. (101a) and (101b) show the change in agreement morphology as a result of NI.<sup>8</sup>

---

<sup>8</sup>The capital H in these examples stands in place of the orthographic apostrophe signaling a glottal stop which is used in the segmented orthography line to be compatible with the LKB software.

- (101) a. *wa'kehninu'*                      *kanakta'*  
 waH-ke-hninu-H                      kanaktaH  
 fact-1SGS.NSGO-buy-PUNC NSG.bed  
 'I bought the/a bed.' [moh] (Baker, 1996:12 via Baker et al., 2005 139)
- b. *wa'kenakthninu'*  
 waH-ke-nakt-hninu-H  
 FACT-1SGS-bed-buy-PUNC  
 'I bought the/a bed.' [moh] (Baker, 1996:12 via Baker et al., 2005 139)

INs in Mapudungun can strand modifiers such as the demonstrative in (102) or the adjective in (103); they can also appear with doubled nouns such as (104).

- (102) *Thikv vyenaktnuhwe"*  
 Thikv v-ye-nakt-nuhweH-H  
 this FUT-FSGS-bed-like-PUNC  
 'She will like this bed.' [moh] (Baker, 1996:308 via Baker et al., 2005:162)
- (103) *Asehtsi vyenaktnuhwe"*  
 Asehtsi v-ye-nakt-nuhweH-H  
 new FUT-FSGS-bed-like-PUNC  
 'She will like the new bed.' [moh] (Baker, 1996:308 via Baker et al., 2005:162)
- (104) *sha'teku rabahbot wa'hvtsyhninu'*                      *rake'niha*  
 shaHteku rabahbot waH-ra-tsy-hninu-H                      rake-Hniha  
 eight      bullhead FACT-MSGS.ZPLO-fish-buy-PUNC my-father  
 'my father fish-bought eight bullheads.' [moh] (Baker, 1996:310 via Baker et al.,  
 2005:149)

Stranding results in an intransitive sentence, evidenced by the agreement affixes only referencing the subject. Doubling, on the other hand, results in a transitive sentence with two-part agreement. Thus, it is ungrammatical for an NI construction with stranding (i.e. a verb inflected with the **strand-mod-lex-rule**) to co-occur with any transitive agreement affixes and likewise for doubling constructions and intransitive agreement. In order to capture these facts, I used the forbids constraints on the NI subpage of the customization system. I made the doubling strategy forbid the intransitive agreement affix (verb-pc3) and the

Stranding strategy forbid the transitive agreement affix (verb-pc2).<sup>9</sup> The corresponding choices are shown below.

```
(105) section=noun-incorporation
      noun-incorp=on
      reduce=on
      red-val=red-intrans
      red-val-intrans-case=red-val-intrans-no
        reduce-forbid1_others=verb-pc2
      promote=on
      promote-obl=on
      ni-predname=benefactive
        promote-forbid1_others=verb-pc3
      double=on
      double-noun=on
        double-forbid1_others=verb-pc3
      strand-mod=on
        strand-forbid1_others=verb-pc2
```

Mohawk also allows oblique arguments promotion as in (106). Two part agreement with the subject and promoted argument is required, and a forbids constraint with verb-pc3 seen above likewise prohibits this type of NI from occurring with intransitive agreement morphology.

```
(106) wa'kheyathvnotsheru'           owira'a
      waH-khey-athvnotsher-u-H       owiraHa
      FACT-1SGS.FSGO-ball-give-PUNC baby
      'I gave the ball to the baby.' [moh] (Baker, 1996:207 via Baker et al., 2005:140)
```

Even with the addition of the forbid constraints, I found that this grammar was over generating negative examples by allowing transitive verbs to be inflected with intransitive morphology without the presence of NI, since the NI position class was optional. Upon inspecting the grammar produced by the customization system, I found that the type `verb-lex` was inheriting from `intrans-agreement-rule-dtr`, a supertype which allows the output of

---

<sup>9</sup>Recall, these morphotactic constraints cannot be defined on the Morphology subpage of the customization system since the user does not directly define a LRT for stranding or doubling—these are created behind the scenes.

a lexical rule to feed another. The problem was that some types of NI are compatible with transitive morphology and other types with intransitive morphology, but all stem from the same position class within verbal inflection. The portion of the customization process which assigns supertypes ‘percolates’ these supertypes to the source. Since transitive verbs inflected by the `reduce-lex-rule` can co-occur with intransitive morphology (with NI), and intransitive verbs always have intransitive morphology, the common `verb-lex` type received `intrans-agreement-rule-dtr` as a supertype. The morphotactic constraints that I defined were sufficient in managing the co-occurrence of agreement morphology with the correct NI types, but in sentences without incorporation, the intermediate types led to undesired behavior. This issue can only be solved by hand-editing the TDL, which I did to improve the coverage of the grammar. I removed `intrans-agreement-rule-dtr` from the definition of `verb-lex`, and added it to the definition of `intransitive-verb-lex`.<sup>10</sup>

The performance of the Mohawk grammar is shown in Table 6.3. All 15 of the grammatical strings parsed, with an average of 1.4 parses per sentence. This ambiguity arises because of the two promotion rules and the ability for demonstratives to modify both the IN and other overt arguments. The value reported for overgeneration is high (50%), but of the three ungrammatical strings that have a parse, there is an alternative, grammatical reading of the sentence. None of the parses reflect the intended ungrammatical readings.

### 6.1.5 *Tongan [ton]*

Tongan is a Austronesian language spoken in Tonga, a group of islands in Oceania. I created a testsuite of seven grammatical and three ungrammatical sentences from Rosen, 1989 and Ball, 2005.

It displays an ergative/absolutive adpositional case marking system, and the word order is fairly flexible, allowing both VSO and VOS orders. For the purpose of creating a small grammar to test NI, I defined the word order as verb-initial in customization to allow for

---

<sup>10</sup>I expect this issue to not be uncommon among users of my library, and have added a validation warning to alert users to the potential of this same problem occurring in their grammars.

both dominant word orders. Verbal inflection is quite limited, with no agreement morphology for core arguments, and TAM information encoded with auxiliary verbs. Some literature on NI in Tongan simply juxtaposes the IN to the verb root (c.f. Massam, 2009, Gerdtts, 2017) while others represent this combination morphologically (c.f. Rosen, 1989). As my library is designed to handle INs as lexical rule types for verbal roots, I adopt the convention of affixing the INs, even when the source text does not. The testsuite for Tongan consists of seven grammatical and two ungrammatical sentences.

Similar to Chukchi, the ergative case marking present on subjects of transitive verbs (107a) becomes absolutive when a noun is incorporated (107b).

- (107) a. *nae haka e he siana a he ika*  
 nae haka e he siana a he ika  
 PAST cook ERG the man ABS the fish  
 ‘The man cooked a fish.’ [ton] (Chung, 1978:152 via Rosen, 1989:311)
- b. *nae hakaika a he siana*  
 nae haka-ika a he siana  
 PAST cook-fish ABS the man  
 ‘The man cooked fish.’ [ton] (Chung, 1978:152 via Rosen, 1989:311)

Tongan also allows the stranding of numerous modifiers, including adjectives (108), coordinated NPs (109), and adpositional phrases (110).

- (108) *nae takita foou a Sione*  
 nae ta-kita foou a Sione  
 PAST hit-guitar new ABS Sione  
 ‘Sione played a new guitar.’ [ton] (Ball, 2005:9)
- (109) *nae tomanioke moe talo a Sione*  
 nae to-manioke moe talo a Sione  
 PAST plant-cassava with taro ABS Sione  
 ‘Sione planted cassava and taro.’ [ton] (Ball, 2005:10)
- (110) *nae fakamaasea i fale a Sione*  
 nae fakamaa-sea i fale a Sione  
 PAST clean-chair in house ABS Sione  
 ‘Sione cleaned chairs in the house.’ [ton] (Ball, 2005:10)

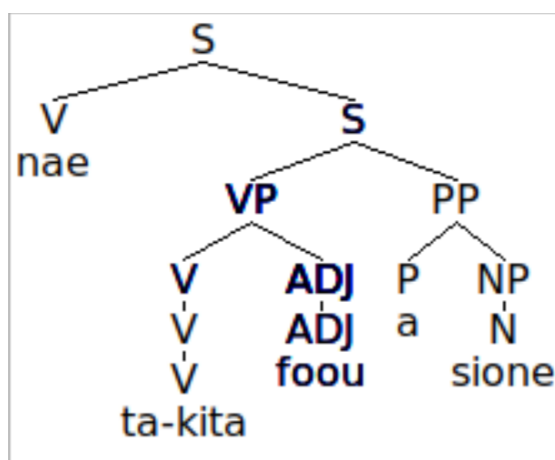


Figure 6.1: Parse tree for sentence (108)

Each of these stranded modifiers goes through the `adj-ni-mod-phrase`.

Unlike the stranded coordinated NP in Southern Tiwa which would not parse, the one for Tongan does. This is because I analyze Tongan as a ‘WITH-language’ in terms of coordination. Drellishak and Bender (Drellishak and Bender, 2005), in defining the Grammar Matrix library for coordination, restrict it to ‘AND-languages,’ and observe that the ‘WITH-language’ coordination strategy is syntactically identical to adjuncts. Thus, in Tongan, the adpositional phrase created by the noun and *moe* ‘with’ can be analyzed as a modifier of the head noun, rather than a true coordination of two NPs. Further, this analysis allows the PP *i fale* ‘in house’ to modify the IN just as the adjective *foou* does. Recall that the HEAD value on the head daughter of `adj-ni-mod-phrase` is constrained to be *+jp* which is the disjunctive type for adjectives and adpositions, allowing it to apply to both the adjective and prepositional phrase. Parse trees for (108) and (109) are shown in Figure 6.1 and Figure 6.2. The second ADJ node in Figure 6.1 and the second PP node in Figure 6.2 both represent the application of the `adj-ni-mod-phrase` non-branching phrase structure rule. The mother node in each case combines the modifier and the inflected verb via the `head-comp` rule.

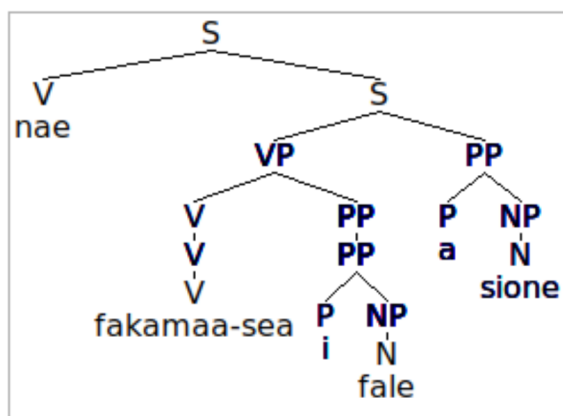


Figure 6.2: Parse tree for sentence (110)

The performance of the Tongan grammar against the testsuite is shown in Table 6.3. 7/7 of the grammatical sentences parsed, with only one parse per sentence. No ungrammatical sentences parsed, indicating that the library was able to fully capture the behavior of NI in Tongan.

Language	Coverage	Overgeneration	Ambiguity
Chukchi	100%	0%	1.43
S. Tiwa	92.3%	0%	1.58
Mapudungun	100%	25%	1
Mohawk	100%	50%	1.4
Tongan	100%	0%	1

Table 6.3: Illustrative language grammar performance

## 6.2 Pseudo-languages

In the process of refining my analysis during implementation with the real, illustrative languages described in the previous section, there were some gaps in the coverage of the tests that I noticed. I crafted six pseudo-languages to address these gaps. The distribution of NI choices and other language features can be seen in Tables 6.4 and 6.5. In general, I focused on making sure I had tested my system with different combinations of NI choices with word order, case marking, argument agreement, and transitivity marking systems. I also created one pseudo language with the three main valence change operations provided by the valence change library (Curtis, 2018). All six of the pseudo-languages include both Promotion NI choices as well as Stranding and Doubling since these options are not mutually exclusive. Each has one of the Reduction NI choices: transitive, intransitive with no case change on the subject, and intransitive with case change on the subject.

Language	Reduction NI			Promotion NI		Modifier NI	
	trans	intrans same	intrans different	promote poss	promote obl	strand	double
PL1	✓			✓	✓	✓	✓
PL2		✓		✓	✓	✓	✓
PL3			✓	✓	✓	✓	✓
PL4	✓			✓	✓	✓	✓
PL5			✓	✓	✓	✓	✓
PL6	✓			✓	✓	✓	✓

Table 6.4: Pseudo-language choices on NI subpage

The choice of what word order or case system the pseudo-language had came from examining the combinations of choices in the illustrative languages considered. For example, the only language with the transitive version Reduction NI, Southern Tiwa, had free word

Language	WO	case system	subj/obj agreement	transitivity marking	valchange ops
PL1	SOV	erg/abs	subj, obj	yes	
PL2	SVO	nom/acc	subj, obj	yes	
PL3	VSO	erg/abs	subj, obj	yes	
PL4	OSV	tripartite	subj+obj	yes	
PL5	OVS	tripartite	subj, obj	yes	
PL6	VOS	(nom/acc), benefactive, instrumental, causative	subj+obj	no	passive, applicative, causative

Table 6.5: Pseudo-language additional choices for testing library interaction

order and, as I analyzed it, an underlying nominative accusative case system for core arguments. I include two pseudo-languages (PL1 and PL3) which have ergative/absolutive case marking and different word orders. In order to keep the number of pseudo-languages manageable, I do not account for every combination across each choice variable, but do strive to diversify the coverage space. Some combinations do not make sense to account for and are not represented. For example, in languages with an ergative/absolutive case system, transitive subjects pattern differently than intransitive subjects. Thus if an ergative/absolutive language behaves intransitively in NI constructions, the case value on the subject should pattern like other intransitive subjects. Therefore I do not have a pseudo-language with both an ergative/absolutive case system and an intransitive Reduction NI strategy where the case value does not change on the subject as this would be inconsistent with intransitive behavior.

Four of the five illustrative languages employed some type of argument agreement inflection on the verb whether this was one morpheme inflected for both the subject and object,

or two separate morphemes for each core argument. All six of the pseudo-languages contain an agreement inflection strategy, with the majority having two separate affixes since this strategy was less common among the illustrative languages. For pseudo-languages with the Reduction NI choice, the object agreement affix agrees with the incorporated noun. For pseudo-languages with the intransitive Reduction NI choices, the object agreement affix does not appear in incorporation constructions.

None of the illustrative languages considered had any transitivity marking inflectional morphology, however I know this to be a feature of some languages. Thus in the pseudo-languages, each have some sort of transitivity inflection. I defined both a transitive and intransitive morpheme. For those pseudo-languages with intransitive Reduction NI, I made sure that these constructions co-occurred with the intransitive morpheme rather than the transitive one.

An example of one of my pseudo-language testsuites for is shown below.

(111) girl-erg cookie-abs bake-TR-3FemSgS-3NeutSg0  
 girl-erg bake-cookie-TR-3FemSgS-3NeutSg0  
 girl-abs smile-IT-3FemSgS  
 \*girl-erg bake-cookie-IT-3FemSgS-3NeutSg0  
 \*girl-abs bake-cookie-TR-3FemSgS  
 \*girl-abs bake-cookie-IT-3FemSgS  
 \*girl-abs bake-cookie-TR-3FemSgS-3NeutSg0  
 \*cookie-abs girl-erg bake-TR-3FemSgS-3NeutSg0  
 \*bake-cookie-TR-3FemSgS-3NeutSg0 girl-abs  
 girl-erg friend-abs bake-cookie-TR-3FemSgS-3MascSg0  
 \*girl-erg friend bake-cookie-TR-3FemSgS-3MascSg0  
 \*girl-erg friend-abs bake-cookie-TR-3FemSgS  
 \*girl-erg friend-abs bake-cookie-TR-3FemSgS-3NeutSg0  
 \*friend-abs girl-erg bake-cookie-TR-3FemSgS-3MascSg0  
 \*friend-abs bake-cookie-TR-3FemSgS-3MascSg0 girl-erg  
 \*bake-cookie-TR-3FemSgS-3MascSg0 girl-erg friend-abs  
 \*bake-cookie-TR-3FemSgS-3MascSg0 friend-abs girl-erg  
 girl-erg yummy bake-cookie-TR-3FemSgS-3NeutSg0  
 \*girl-erg bake-cookie-TR-3FemSgS-3NeutSg0 yummy  
 girl-erg snickerdoodle-abs bake-cookie-TR-3FemSgS-3NeutSg0  
 \*girl-erg bake-cookie-TR-3FemSgS-3NeutSg0 snickerdoodle-abs  
 \*girl-erg snickerdoodle bake-cookie-TR-3FemSgS-3NeutSg0  
 \*girl-abs snickerdoodle bake-cookie-TR-3FemSgS-3NeutSg0

```
*girl-erg snickerdoodle-abs bake-cookie-IT-3FemSgS-3NeutSgO
*girl-erg snickerdoodle-abs bake-cookie-TR-3FemSgS-3MascSgO
*snickerdoodle-abs girl-erg bake-cookie-TR-3FemSgS-3NeutSgO
```

Each line is a string that is either grammatical or ungrammatical (marked with an asterisk \*) in the language. In constructing these pseudo-languages, I use a limited English vocabulary for nouns, verbs, and adjectives, and typical glossing conventions for verbal and nominal morphology. This is for relative ease of comprehension when reading the strings. PL1 has transitive Reduction NI, SOV word order, ergative absolutive case marking, transitivity marking, and separate agreement affixes for subject and object for which it agrees in terms of person, gender, and number. The string in (112) is a grammatical sentence in the testsuite for PL1.

(112) *girl-erg bake-cookie-TR-3FemSgS-3NeutSgO*

The noun *cookie* is incorporated into the verb *bake* and no promotion, stranding or doubling occurs. Therefore, the verb is further inflected with the transitive morpheme, and the object agreement affix agrees with the IN.<sup>11</sup> Each ungrammatical strings differ from a grammatical one with one change. For example, the string in (113) is ungrammatical due to the intransitive morpheme and the string in (114) is ungrammatical because the subject receives absolutive case marking.

(113) *\*girl-erg bake-cookie-IT-3FemSgS-3NeutSgO*

(114) *\*girl-abs bake-cookie-TR-3FemSgS-3NeutSgO*

Selected elements of the choices file used to generate the grammar for PL1 are given below. Elements not included define the word order, case system, number and gender definitions as well as subject marking inflection.

(115) `section=arg-opt`  
`obj-drop=obj-drop-all`

---

<sup>11</sup>In each of the pseudo-languages, I adopted the convention of specifying feminine gender on the noun *girl*, masculine gender on *friend*, and neuter gender on *cookie* and *snickerdoodle*.

obj-mark-drop=obj-mark-drop-req  
 obj-mark-no-drop=obj-mark-no-drop-req

section=noun-incorporation  
 noun-incorp=on  
 reduce=on  
 red-val=red-trans  
 promote=on  
 promote-poss=on  
 promote-obl=on  
 ni-predname=benefactive  
 double=on  
 double-noun=on  
 strand-mod=on

section=morphology  
 verb-pc1\_name=noun\_incorporation  
 verb-pc1\_order=suffix  
 verb-pc1\_inputs=tverb  
 verb-pc1\_is-lrt1\_name=incorporated\_noun  
 verb-pc1\_is-lrt1\_feat1\_name=number  
 verb-pc1\_is-lrt1\_feat1\_value=sing  
 verb-pc1\_is-lrt1\_feat1\_head=in  
 verb-pc1\_is-lrt1\_feat2\_name=person  
 verb-pc1\_is-lrt1\_feat2\_value=3rd  
 verb-pc1\_is-lrt1\_feat2\_head=in  
 verb-pc1\_is-lrt1\_feat3\_name=gender  
 verb-pc1\_is-lrt1\_feat3\_value=neuter  
 verb-pc1\_is-lrt1\_feat3\_head=in  
 verb-pc1\_is-lrt1\_lri1\_inflecting=yes  
 verb-pc1\_is-lrt1\_lri1\_orth=-cookie  
 verb-pc1\_is-lrt1\_lri1\_pred=\_cookie\_n\_rel  
 verb-pc2\_name=trans  
 verb-pc2\_obligatory=on  
 verb-pc2\_order=suffix  
 verb-pc2\_inputs=tverb, verb-pc1  
 verb-pc2\_lrt1\_name=TR  
 verb-pc2\_lrt1\_lri1\_inflecting=yes  
 verb-pc2\_lrt1\_lri1\_orth=-TR  
 verb-pc3\_name=intrans  
 verb-pc3\_obligatory=on  
 verb-pc3\_order=suffix  
 verb-pc3\_inputs=iverb  
 verb-pc3\_lrt1\_name=ITR

```

    verb-pc3_lrt1_lri1_inflecting=yes
    verb-pc3_lrt1_lri1_orth=-IT

verb-pc6_name=trans-obj
verb-pc6_obligatory=on
verb-pc6_order=suffix
verb-pc6_inputs=verb-pc4
  verb-pc6_lrt1_name=3NeutSg0
    verb-pc6_lrt1_feat1_name=number
    verb-pc6_lrt1_feat1_value=sing
    verb-pc6_lrt1_feat1_head=obj
    verb-pc6_lrt1_feat2_name=gender
    verb-pc6_lrt1_feat2_value=neuter
    verb-pc6_lrt1_feat2_head=obj
    verb-pc6_lrt1_feat3_name=person
    verb-pc6_lrt1_feat3_value=3rd
    verb-pc6_lrt1_feat3_head=obj
    verb-pc6_lrt1_lri1_inflecting=yes
    verb-pc6_lrt1_lri1_orth=-3NeutSg0

```

Since this pseudo-language has transitive Reduction NI,<sup>12</sup> object dropping must be enabled on the argument optionality subpage (`section=arg-opt`).<sup>13</sup> From the Noun Incorporation subpage (`section=noun-incorporation`), the `red-val` key has a value of `red-trans` to indicate transitive behavior in Reduction NI constructions. All other NI options are marked as ‘on.’

On the morphology subpage (`section=morphology`), I defined a PC for noun incorporation with one IS-LRT. Again owing to the transitive nature of Reduction NI and the fact that object agreement morphology targets the IN, I define number, person, and gender values (singular, 3rd, neuter) for this IS-LRT, and give it an orthographic form *-cookie* with a predicate value `_cookie_n_rel`.

The final thing of note is `verb-pc6` which specifies agreement features for the object. In `verb-pc6_lrt1` these features require the object to be 3rd person, singular in number,

---

<sup>12</sup>For users wondering if the language they are modeling has transitive or intransitive Reduction NI, the Noun Incorporation subpage asks them to consider markers of transitive behavior such as object agreement markers or other transitive morphology.

<sup>13</sup>As explained in some the Implementation chapter, transitive Reduction NI keeps the IN on the verb’s COMPS lists and constrains it to be [OPT –].

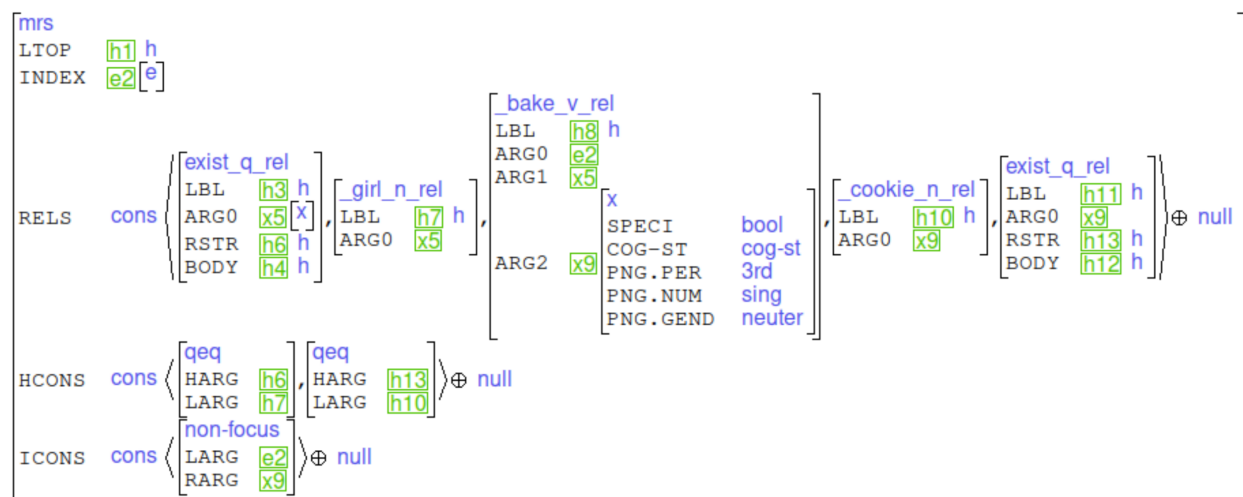


Figure 6.3: Semantic representation for pseudo-language sentence (112)

and have neuter gender. These features are specified on the object of the verb (**head=obj**), which, in NI constructions, will reference the IN which remains on the verb’s COMPS list.

As with illustrative languages, for each sentence in the testsuite, I manually inspect the tree structure and MRS representation to ensure that the grammar behaves as intended. An example MRS for the grammatical example sentence above is given in figure 6.3. The semantics correctly captures the syntactic role of *cookie* as the object of the verb as well as the PNG features of the IN.

The results for each pseudo-language grammar are shown in table 6.6. For PL1, 1/20 ungrammatical strings parsed. Upon inspection, the tree built does represent a valid interpretation of the string, just not the one envisioned when constructing negative examples. The same is true for the 5% overgeneration reported for PL4. There were an average of 1.69 parses per grammatical string across all pseudo-language. This ambiguity is expected based on the symmetry of the promoted possessor and promoted oblique argument constructions. There is also the ability to interpret a doubled noun as a promoted object, which results in

Language	Coverage	Overgeneration	Ambiguity
PL1	100%	5%	1.5
PL2	100%	0%	1.57
PL3	100%	0%	1.57
PL4	100%	5%	1.43
PL5	100%	0%	1.57
PL6	100%	0%	2.5

Table 6.6: Pseudo-language grammar performance

multiple parses for one sentence. Across the other five pseudo-languages, the same patterns held true. All resulted in 100% of grammatical strings parsing, with ambiguity between Promotion NI types and doubling accounting for the minimal (average 1.69 parses per sentence) spurious ambiguity.

### 6.3 *Held-out Languages*

After refining my analysis and implementation to maximize the performance of the illustrative and pseudo-language grammars over their constructed testsuites, I froze development to validate my library against unseen and unconsidered data. I selected five languages described to contain noun incorporation which had sufficient data and description to construct a test-suite. It is important to note that I constructed these testsuites independent of my analysis to most accurately reflect how NI actually appears in the languages. These languages are Apurinã [apu], Bribi [bzd], Inuktitut [iku], Moloko [mlw], and Yaqui [yaq]. The types of NI found in each language are summarized in Table 6.7, and additional information about each’s typology is shown in Table 6.8. In this section, I describe the characteristics of these languages, including their manifestations of NI. I discuss the performance of my library in capturing the facts about NI for each language and discuss areas where my analysis falls

short.

Language	Reduction NI			Promotion NI		Modifier NI	
	trans	intrans same	intrans different	promote poss	promote obl	strand	double
Apurinã	✓			✓		✓	✓
Bribri			✓	✓	✓		
Inuktitut			✓		✓	✓	✓
Moloko		✓					
Yaqui		✓					

Table 6.7: Held-out language choices on NI subpage

### 6.3.1 *Apurinã [apu]*

Apurinã is an Indigenous Amazonian language spoken in Brazil. It does not employ a case marking system for core arguments, though oblique arguments are marked with the goal case. It has two agreement markers for the subject and object in terms of person and number. The testsuite for Apurinã comes from ten grammatical sentences adapted from Facundes (2000) and 13 hand-constructed ungrammatical sentences. Pronominal cross referencing of the core arguments is not allowed when overt nouns proceed the verb, but optional when they follow. The Grammar Matrix is currently unable to account for different relationships between argument optionality and agreement marking based on word order, so I edited the grammatical sentences found in the reference text to move all nouns post-verbally, and allow argument dropping in all cases.

When the sole object of a verb is incorporated, transitive agreement morphology remains. The sentence in (116) shows the object agreement affix agreeing with the IN. All NI constructions require the presence of the verbalizing affix *-ta*.

Language	WO	case system	subj/obj agreement	transitivity	valchange ops
Apurinã	VSO	(nom/acc), goal	subj, obj	no	
Bribri	free	erg/abs, intrs, com, loc, ben	none	imperfective	obj-add
Inuktitut	SOV	erg/abs, mod, allat	subj+obj	no	passive, antipassive
Moloko	SVO	(nom/acc), loc, dat	subj, ind obj	no	obj-add
Yaqui	SOV	nom/acc	none	yes	passive

Table 6.8: Held-out language additional choices for testing library interaction

- (116) *nutakapetaru*  
 nu-taka-pe-ta-ru  
 1SGS-put-pulp-VBLZ-3SGO  
 ‘I put the paste object.’ [apu] (Facundes, 2000:306)

Incorporation of possessed nouns can either lead to possessor raising as in (117a) or doubling of the IN as in (117b).<sup>14</sup>

- (117) a. *usonakapetaka ata komuru*  
 usonaka-pe-ta ata komuru  
 dry-pulp.of-VBLZ 1PL manioc  
 ‘We put the manioc pulp to dry.’ [apu] (Facundes, 2000:169)
- b. *usonakapetaka ata komuru pe*  
 usonaka-pe-ta ata komuru pe  
 dry-pulp-VBLZ 1PL manioc pulp.of

---

<sup>14</sup>This sentence originally includes an additional affix which is glossed as both a causative and an intensifier in the same paragraph of Facundes (2000:169). In the causative reading, this would first involve the incorporation of a doubled subject which is demoted to object when the causative morpheme adds a subject. Subject incorporation is not supported by my library, so modeling the sentence with a causative morpheme would not be permitted. Instead, I assume the intensifier reading, but remove this morpheme as it is not yet covered by other Matrix libraries.

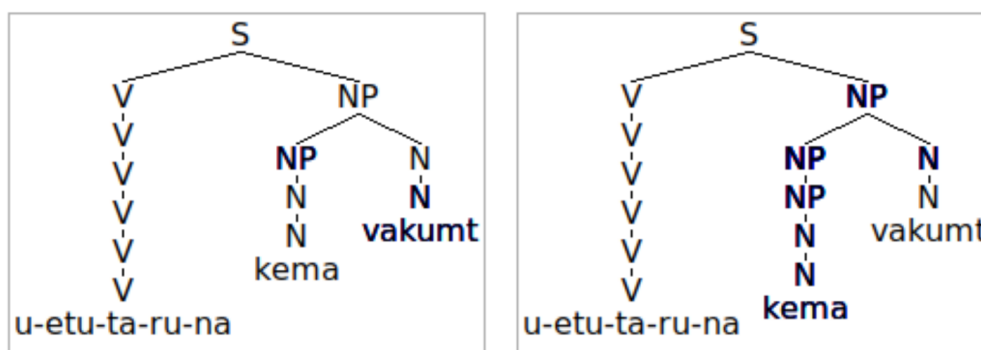


Figure 6.4: Two parses generated for sentence (118)

‘We put the manioc pulp to dry.’ [apu] (Facundes, 2000:307)

The performance of the Apurinã grammar is summarized in Table 6.9. All ten of the grammatical sentences parsed, but there were an average of 2.1 parses per sentence, showing some ambiguity. When evaluating these ambiguous parses, I found an issue with my library. Possessive constructions are built with a phrase structure rule which combines the possessor and the possessum. I did not account for this type of phrase structure when constraining the NCORP-MOD feature I defined for stranded modifiers and doubled nouns. Figure 6.4 shows the two parses generated for the sentence in (118).

- (118) *uetutaruna*                      *kema vakumt*  
 u-etu-ta-ru-na                      kema vakumt  
 3M-see-VBLZ-3M.O-3PL tapir footprint.of  
 ‘They saw the tapir’s footprint.’ [apu] (Facundes, 2000:303)

Both parse trees show an NP node representing the structure licensed by the `noun-poss-1` rule for forming a possessive noun phrase. The tree structure on the right shows that the non-head daughter, the possessor, is not constrained to be [NCORP-MOD –] as it should be. This mistake accounts for 6/11 of the ambiguous parses.

The other five ambiguous parses are due to the unavoidable ambiguity of the `double-noun`

-rule and promote-poss-rule. This ambiguity was seen also in the illustrative languages, and is expected.

3/13 of the ungrammatical sentences in the Apurinã testsuite parsed. In each case, there is a valid, grammatical interpretation of the input distinct from the intended ungrammatical reading. Thus, the grammar correctly disallows all of the ungrammatical strings as intended.

### 6.3.2 Bribri [bzd]

Bribri is a Chibchan language spoken in Costa Rica. It has an ergative/absolute case marking system for core arguments, as well as a set of oblique argument case markers including instrumental, comitative, locative, and benefactive. Case marking is handled adpositionally except for the absolute case where it is unmarked.<sup>15</sup> The testsuite for Bribri consists of 20 grammatical and 14 ungrammatical sentences. The grammatical sentences come from Pacchiarotti and Zúñiga, 2025, and I constructed the ungrammatical sentences.

There is no agreement morphology for arguments on the verb, but there is a mandatory TAM morpheme. All verbs, intransitive and transitive, must be inflected for tense or aspect. For the imperfective, there is both a transitive and intransitive version as shown in (119) and (120).

- (119) *sah tsaku*  
 sah-abs tsak-u  
 1PL.excl-ABS eat.INTR-IPFV.INTR  
 ‘We are eating.’ [bzd] (Pacchiarotti and Zúñiga, 2025:486)

- (120) *sah tu kutsitsaka kate*  
 sah tu kutsi-tsaka-abs kat-e  
 1PL.EXCL ERG pig-meat-ABS eat.TR-IPFV.TR  
 ‘We are eating pork meat.’ [bzd] (Pacchiarotti and Zúñiga, 2025:486)

---

<sup>15</sup>For simplicity while modeling this language with the customization system, I inserted a null affix *-abs* on unmarked nouns with absolute case. The customization system is unable to properly handle null case marking adpositions, but does support mixed marking strategies (Drellishak, 2009:58). Therefore while ergative case is overtly marked with an adposition, I handled the null absolute case marker as an affix.



In addition to promoted possessors, oblique arguments can also be promoted to object position as in (125). Here, the tongue belongs to the more agent-like argument, the rabbit.<sup>16</sup>

- (125) *sawi tu iadaada apakudje*  
 sawi tu i-adaada-abs apaku-dj-e  
 rabbit ERG 3SG-cub-ABS body.tongue-drink.TR-IPFV.TR  
 ‘The rabbit is licking its cubs.’ [bzd] (Pacchiarotti and Zúñiga, 2025:499)

The performance of the Bribri grammar is shown in Table 6.9. 20/20 grammatical strings parsed, and only 1/14 ungrammatical strings incorrectly parsed. There was an average of 3.7 parses per grammatical sentence, which was the highest level of ambiguity among all languages reported in this thesis. The reason for such high ambiguity is mostly due to adverbs in the language. Adverbs in the customization system are not yet fully supported, so there is not a way outside of TDL editing to constrain at what level (V, VP, or S) the adverb applies. In particular, one sentence with two adverbs received 43 parses, heavily skewing the average reported. The only other source of ambiguity in this testsuite is the fact that both Promotion NI types are relevant to the language. Any time that one applies, a duplicate parse using the other rule appears as expected. The one ungrammatical parse reported has a different interpretation of the input string which is grammatical. In all, the grammar created with the addition of my library correctly handles every grammatical sentence, no ungrammatical meanings, and is not ambiguous apart from known issues.

### 6.3.3 *Inuktitut [iku]*

Inuktitut, also known as Eastern Canadian Inuit, is an Eskaleut language spoken in Canada. I created a testsuite for the language consisting of 27 grammatical and 27 ungrammatical sentences. The grammatical and some ungrammatical examples come from Yuan (2022), which describes NI in the language.

---

<sup>16</sup>Originally, the sentence contained a more complex promoted noun phrase where *i-adaada* was further modified by the adjective *buk* ‘two.’ However, the lexical entry for adjectives requires that the element they modify have a non-empty SPR list. Since this noun is possessed, it has an empty SPR list. Resolving this issue would involve TDL editing which is beyond the scope of validating my library only using customization system-generated grammars.

Inuktitut has a standard ergative absolutive pattern for case marking in transitive sentences (126a). However, two valence changing operations impact case marking. In antipassive constructions, the subject is marked with absolutive case while the object is marked with the oblique modalis case (126b). In passive constructions, the patient-like subject is marked with absolutive case, while the agent-like object is marked with the allative case (127).

(126) a. *Taivitiup igalaaq suraktaa*  
 Taiviti-up igalaaq surak-taa  
 David-ERG window.ABS break-3SG.S/3SG.O  
 ‘David broke the window.’ [iku] (Yuan, 2025:517)

b. *Taiviti igalaaqmik suraksijuq*  
 Taiviti igalaaq-mik surak-si-juq  
 David.ABS window-MOD break-AP-3SG.S  
 ‘David broke the window.’ [iku] (Yuan, 2025:517)

(127) *asi angajuqanut suatanngitjut*  
 asi angajuqa-nut sua-ta-nngit-jut  
 other parent-ALLAT scold-PASS.PART-NEG-3PL.S  
 ‘The others were not scolded by their parents.’ [iku] (Yuan, 2025:525)

When NI occurs, the sentence often patterns like the antipassive construction above. The subject, when overt, takes the unmarked absolutive case. Any stranded modifiers of the IN are marked with the modalis case. This can be seen in (128).

(128) *Ulak piunki ujamiliujuq*  
 Ulak piu-mik ujami-liu-juq  
 Ulak.ABS beautiful-MOD necklace-make-3SG.S  
 ‘Ulak made a beautiful necklace.’ [iku] (Yuan, 2025:518)

However, the antipassive pattern is not the only behavior for NI. Consider the pair of sentences in (129a) and (129b) below.

(129) a. *Taiviti sivalaartuqrumajuq*  
 Taiviti sivalaar-tuq-ruma-juq  
 David.ABS cookie-consume-want-3SG.S  
 ‘David wants to eat the cookie/cookies.’ [iku] (Yuan, 2025:533)

- b. *Taivitiup sivalaartuqrumajangit*  
 Taiviti-up sivalaar-tuq-ruma-jangit  
 David-ERG biscuit-consume-want-3SG.S/3PL.O  
 ‘David wants to eat these cookies.’ [iku] (Yuan, 2025:533)

Syntactically, the only difference between the two is the case marking on the subject, and the agreement affix on the verb. Semantically, there is a minor distinction in the translations, where the transitive reading of *sivalaar* ‘cookies’ is more concrete. When designing my analysis, I made the claim that no language would have both transitive and intransitive Reduction NI. While these data might challenge that assumption, it is also possible to interpret them another way. While (129a) seems to clearly be an example of Reduction NI, (129b) can be analyzed as a doubling construction where the doubled noun is dropped. This aligns with the transitive morphology on the verb, and the semantic distinction between the two.

As implemented, the **double-noun-lex-rule** requires that the doubled noun be overt ([OPT –]). This restriction means that the ergative-pattern NI sentences like (129b) do not parse as intended. This represents a limitation of my library. I made the assumption that promoted and doubled objects would always be overt. It would have been better to allow the user to specify whether such arguments must appear or if they can be dropped.

Verbs with incorporated nouns can also be passivized as in (130). Like in standard passive constructions, the patient-like argument is marked with the allative case.

- (130) *nuliaqtanut aasivartuqtajuq*  
 nuliaqta-nut aasivar-tuq-ta-juq  
 mate-ALLAT spider-consume-PASS.PART-3SG.S  
 ‘The spider is being eaten by its mate.’ [iku] (Yuan, 2025:510)

There are two ways to analyze this construction. Either passivization occurs first where *aasivar* ‘spider’ is promoted to subject and *nuliaqta* ‘mate’ is demoted to object. In this case, the incorporation of *aasivar* ‘spider’ is a matter of subject incorporation which is not supported by this library. The other analysis is that incorporation happens before passivization. The IN remains incorporated, and *nuliaqta* ‘mate’ is demoted from subject

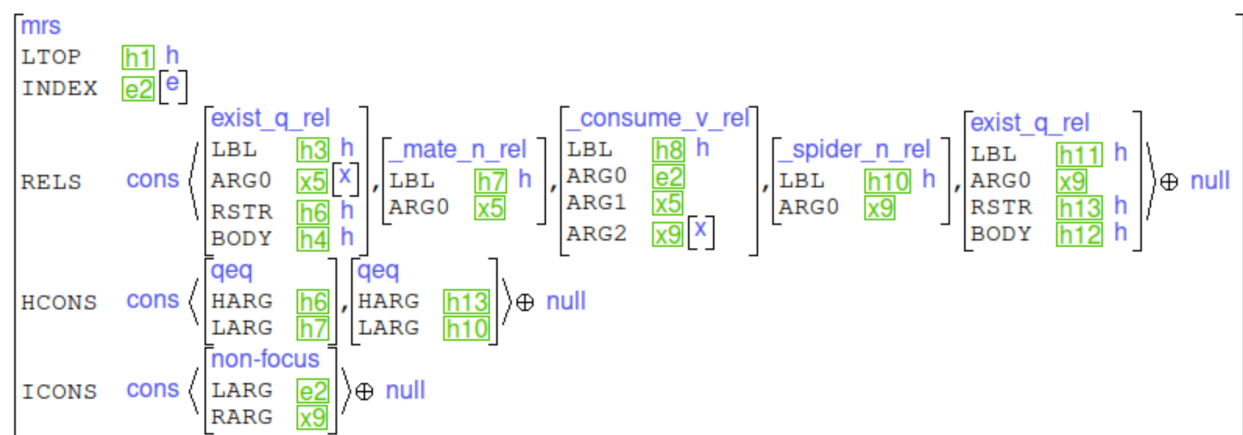


Figure 6.5: Semantic representation for sentence (130)

to object. The problem with this analysis is that at the point passivization occurs, the verb is intransitive since the `reduce-lex-rule` ‘empties’ the verb’s COMPS list. The valence change library does not support a subject demotion operation without object promotion. It also does not support passive operations on intransitive targets.

Due to these limitations, the above sentence does not parse as intended.<sup>17</sup> However, the grammar does produce a parse which creates the intended MRS representation. This is achieved with the application of the `strand-mod-lex-rule`. This lexical rule references some element on the verb’s COMPS list (intended to be the stranded modifier), and passivization promotes this element to the verb’s SUBJ list. The passive lexical rule, however, underspecifies some information, including the specification from `strand-mod-lex-rule` that this element be [OPT –]. Thus, the `decl-head-obt-subj` rule is able to apply, dropping the subject. Though this is not the intended inflectional path of the verb, it does produced the desired semantic representation which can be seen in Figure (6.5).

Another feature of passivization of incorporated constructions is the fact that the intran-

<sup>17</sup>It might be possible to take this grammar fragment and edit the TDL of the input to the passive lexical rule to allow intransitive targets to apply there.

sitive subject agreement affix on the verb agrees in number and person with the incorporated noun as seen in examples (131a) and (131b).

- (131) a. *ujamiliutajuq*  
 ujami-liu-ta-juq  
 necklace-make-PASS.PART-3SG.S  
 ‘The necklace is being made.’ [iku] (Yuan, 2025:525)
- b. *ujamiliutajut*  
 ujami-liu-ta-jut  
 necklace-make-PASS.PART-3PL.S  
 ‘The necklaces are being made.’ [iku] (Yuan, 2025:526)

Parses for both of these sentences are made possible by the same combination of **strand-mod-lex-rule** and subject dropping. An additional benefit of this interaction is the ability to identify the person and number features specified by the agreement affix on the IN. For example, Figure (6.6) shows the MRS for (131b), including the PNG features identified with the IN.

Incorporated nouns in passive constructions are able to actually strand modifiers outside of the verb as in (132).

- (132) *una aasivartuqtajuq*  
 una aasivar-tuq-ta-juq  
 DEM.ABS spider-consume-PASS.PART-3S.S  
 ‘This spider was being eaten [iku] (Yuan, 2025:526)

This sentence does not parse, however, due to the customization of the passive lexical rule. I defined the rule so that the subject is marked with absolutive case and the object is marked with allative case. Because of this, the subject is constrained to have a HEAD value of *noun* which is incompatible with the stranded adjective.<sup>18</sup>

The results for the Inuktitut grammar are given in Table 6.9. 21/27 grammatical sentences parsed, including sentences which resulted in the correct semantic representation by means of the unintended structure. Those that did not have been discussed in this section.

---

<sup>18</sup>Removing the case specification on the subject increases overgeneration and ambiguity.

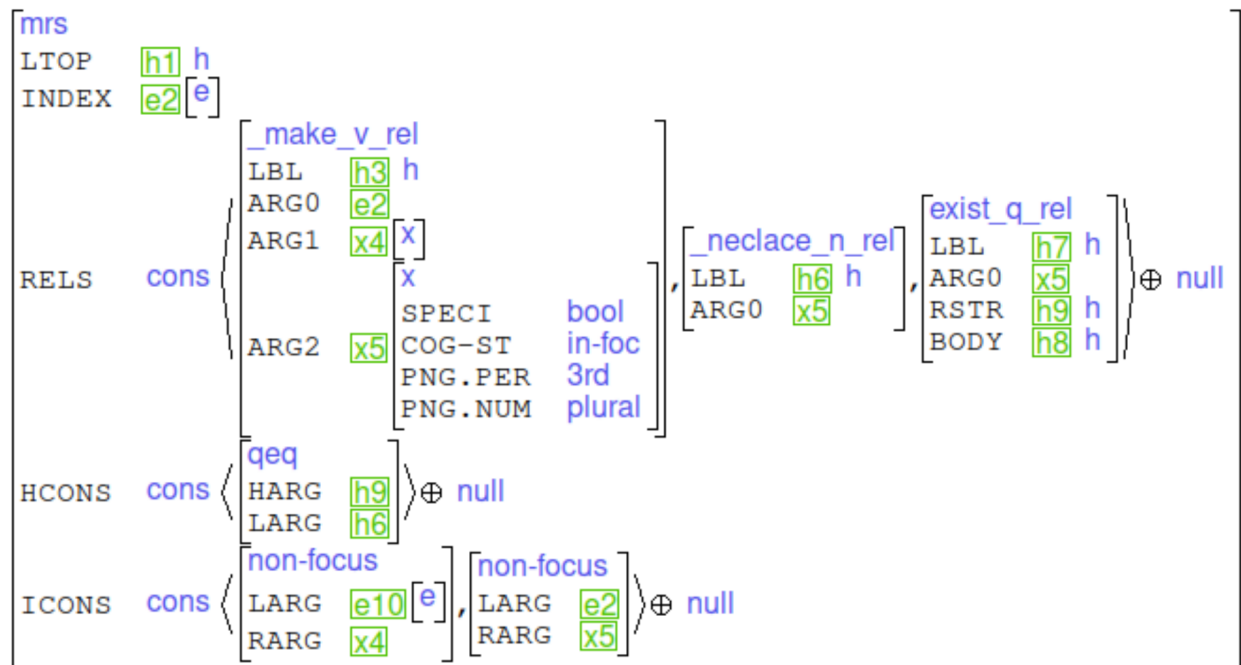


Figure 6.6: Semantic representation for sentence (131b)

To summarize these observations, I was not able to fully account for both the ergative and antipassive patterns of NI due to my library being too strict. I only allow grammars to have either an intransitive or transitive Reduction NI type, and I do not allow users to be able to specify that doubled or promoted arguments can optionally be dropped. These restrictions were made based on observations of languages with NI in the literature, but do not hold up with the facts of Inuktitut. There was an average of 1.62 parses per sentence. This ambiguity arises unavoidably from the enabling of argument optionality, and under-specification on elements promoted and demoted by the passive lexical rule. One such under-specification is the value of NCORP-MOD, which should be specified as negative on all element which the valence change library references. 9/27 ungrammatical sentences received a parse. Standard Doubling is ungrammatical in Inuktitut, as seen in (133).

- (133) *\*tuktu niqituqrumapara*  
 tuktu niqi-tuq-ruma-para  
 caribou meat-consume-want-1S.S/3S.O  
 Intended: ‘I want to meat-eat this caribou meat.’ (Yuan, 2025:536)

I enabled NI doubling in customization to get closer to a representation of the ergative pattern of NI for sentences such as (129b). If I were to un-do this choice, over-generation would not improve as the object can also be realized as a promoted oblique argument.

Like Mohawk [moh], the Inuktitut grammar also has the problem of interpreting transitive verbs as daughter types of intransitive morphology, since transitive verbs inflected with incorporated nouns do pattern intransitively. The other ambiguity comes from the passive rule under-specifying features on the promoted subject and demoted object, and known ambiguities between promotion and doubling rules.

#### 6.3.4 *Moloko [mlw]*

Moloko is a part of the Afro-Asiatic language family, and is spoken in Cameroon. It is described to have NI, but the literature orthographically represents the verb and the IN as separate words. Since my grammar depends on the IN being affixed to a verb root, I



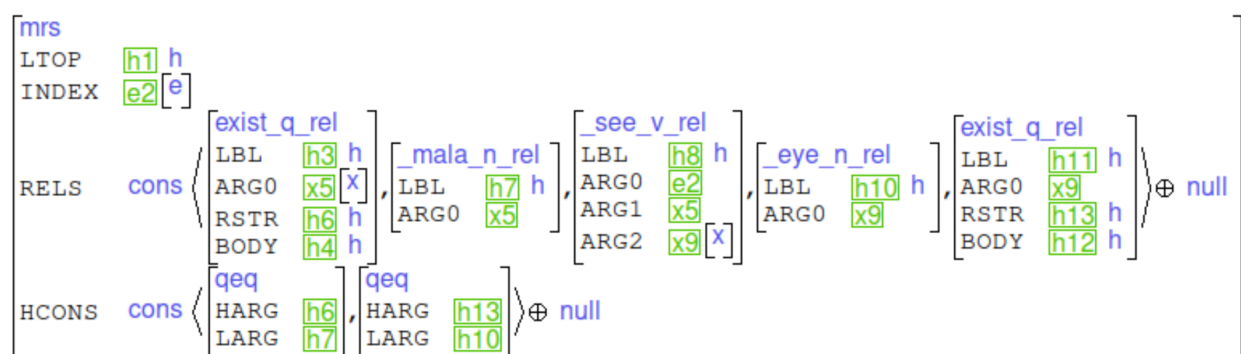


Figure 6.7: Semantic representation for sentence (135)

The Grammar Matrix customization system does not yet support ditransitives, but I was able to model an object-adding valence changing rule to account for such oblique arguments. In order to capture the agreement information, I specified that the feature values associated with this lexical rule were to be marked on the second object.

I chose to analyze constructions such as those in (136) as applications of Reduction NI with valence change operations as opposed to Promotion NI. This is due to the fact that the arguments such as *Hawa* appear with indirect object agreement affixes. Verbs in Moloko can be inflected for both direct and indirect object agreement at the same time as shown in (137). The Promotion NI analysis is not appropriate, then, as it treats the promoted oblique argument as the direct object.

- (137) *akadawnava*  
 a-kad-aw-na-va  
 3SG+PFV-club-1SG.IO-3SG.DO-PRF  
 ‘He/she has killed it for me.’ [mlw] (Friesen, 2017:203)

The Moloko grammar performed the best among the held-out languages, with 10/10 grammatical sentences receiving correct parses, and 0/10 ungrammatical sentences receiving any parse. The minimal (1.2 parses per sentence) ambiguity arose due to the fact that the

object addition valence change rule was non-inflecting and object optionality was enabled. Twice, a parse tree was constructed where a second object was added due to the valence changing lexical rule, only to be dropped by the `basic-head-opt-comp` rule.<sup>20</sup> No ambiguity arose due to my library.

### 6.3.5 Yaqui [yaq]

Yaqui is a Uto-Aztecan language spoken in Mexico. It has nominative/accusative case marking inflection for core arguments and modifiers, transitive and intransitive morphology, and allows NI constructions to be passivized. The testsuite for Yaqui consists of five grammatical and eight ungrammatical sentences adapted from (Haugen, 2008).

In Yaqui, subjects are unmarked for nominative case, and objects are inflected with an accusative case marker. When an adjective modifies an argument, it is inflected for the same case. This can be seen in (138). Notice also the transitive inflection on the verb.

- (138) *aapo maasota bweuuk peutak*  
 aapo maaso-ta bweuu-k peu-ta-k  
 3SG deer-ACC big-ACC butcher-TR-PERF  
 ‘He butchered a big deer.’ [yaq] (Jelinek, 2014:213 via Haugen, 2008:118)

When nouns are incorporated, as in (139), the case marking on the subject remains nominative (unmarked), and the verb is inflected with the intransitive morpheme.

- (139) *aapo maasopeuten*  
 aapo maaso-peu-te-n  
 3SG deer-butcher-INTR-PAST  
 ‘He was deer butchering’ [yaq] (Jelinek, 2014: 213 via Haugen, 2008:118)

NI in Yaqui is limited to simple Reduction. The stranding of modifiers is forbidden (140), and while adnominally possessed nouns may occur as the object of a verb (141), they cannot be incorporated to promote the possessor (142).

---

<sup>20</sup>I would not necessarily expect the agreement marking to be different for these constructions since indirect object agreement is not obligatory.

- (140) *\*aapo bweuuk maasopeuten*  
 aapo bweuu-k maaso-peu-te-n  
 3SG big-ACC deer-butcher-INTR-PAST  
 [\*‘He was [big deer]-butchering’] or [\*‘He was deer-butchering a big one’] [yaq] (Jelinek, 2014: 213 via Haugen, 2008:118)
- (141) *Peo Huanta maaso peuta*  
 Peo Huan-ta maaso peu-ta  
 Peo Huan-POSS deer butcher-TR  
 ‘Peo is butchering Huan’s deer’ [yaq] (Haugen, 2008:153)
- (142) *\*Peo Huanta maasopeute*  
 Peo Huan-ta maaso-peu-te  
 Peo Huan-POSS deer-butcher-INTR  
 ‘Peo is butchering Huan’s deer’ [yaq] (Haugen, 2008:153)

A verb in Yaqui can incorporate a noun and be passivized in the same clause as in (143).

- (143) *Maasopeutewak*  
 Maaso-peu-te-wa-k  
 deer-butcher-INTR-PASS-PERF  
 ‘Deer-butchering occurred.’ [yaq] (Haugen, 2008:153)

Since the IN occurs before the verb root and the passive morpheme occurs after, I had to decide for which the verb was inflected first. If passivization occurred first, then incorporation would target the promoted subject, which is not supported by my library. I chose then to have incorporation feed passivization. However, since Yaqui has an intransitive type of Reduction NI, the verb’s COMPS list is empty at the time the passive morpheme applies. Therefore, instead of using the proto-typical combination of subject-demotion and object-promotion operations, I defined a subject-removing valence-changing operation.<sup>21</sup> This lexical rule correctly parsed the sentence in (143).

---

<sup>21</sup>I was able to model the interaction of passives and NI in the grammar for Yaqui but not for Inuktitut because there was no overt subject in the passive incorporated construction in (143). The limitation described for Inuktitut was the inability to define a subject-demotion rule without a corresponding object-promotion rule. For the Yaqui grammar, I was not restricted to demoting the subject, I could remove it altogether.

Language	Coverage	Overgeneration	Ambiguity
Apurinã	100%	23.1%	2.1
Bribri	100%	7.1%	2.4
Inuktitut	77.8%	33.3%	1.62
Moloko	100%	0%	1.2
Yaqui	100%	12.5%	1

Table 6.9: Held-out language grammar performance

The results of the Yaqui grammar over the created testsuite are shown in Table 6.9. 5/5 grammatical sentences parsed with a perfect one parse-per-sentence ratio. Of the eight ungrammatical sentences, only one received a parse. This is a transitive sentence with intransitive verbal inflection. It was able to parse due to the same problem as what happened with Mohawk [moh] above. The definition for all verbs, `verb-lex` inherits from the daughter type for intransitive inflection because transitive verbs with incorporated nouns are inflected with intransitive morphology. Thus, the grammar as produced allows any transitive verb to take intransitive morphology which is not the desired behavior. The solution to this issue involves hand-editing the TDL to move the `intrans-rule-dtr` supertype to the definition of `intransitive-verb-lex` instead.

### 6.3.6 Summary

In this section, I described how I used the system I built to create grammars for five unconsidered natural languages, with particular focus on manifestations of noun incorporation and its interaction with other linguistic phenomena such as agreement and valence changing operations. Table 6.9 shows the performance of the grammars for each of these held-out languages for three different metrics. In 4/5 of the grammars for held-out languages, I was able to achieve 100% coverage, meaning that all of the grammatical sentences in the testsuites

received a parse. For the Inuktitut grammar, the lower coverage was a result of my library being too strict, and a failed interaction with the valence change library. To improve the coverage, the library should increase the ability to customize the rules of my analysis, for example the ability to state whether or not doubled or promoted arguments can be dropped. However, the number reported here is also generous as it includes the parses which result in the correct MRS but were built with an unintended syntactic structure.

In terms of overgeneration, the numbers here are harsher. They include sentences which were hand crafted to test for unintended behavior, and whose parses do not reflect the intended ungrammatical readings. Taking this into consideration, the true overgeneration averaged across all held-out grammars is 4.72% (compared to the 15.2% reported).

Across the held-out grammars, there was expected and unexpected ambiguity. I expect languages with both Promotion NI types to have semantic ambiguity, and languages with Promotion NI and doubling to have structural ambiguity. Instances of ambiguity that came to light with the held-out languages were due to the NCORP-MOD feature being unconstrained in possessive phrases and on arguments targeted by valence change operations. Overall, the rates of ambiguous parses are fairly low as are true cases of overgeneration. The coverage is also quite high, and can be improved by relatively simple modifications to the library.

#### **6.4 Summary**

In this chapter, I presented my work evaluating the library I made for modeling noun incorporation in the world's languages. I first showed the real 6.1 and constructed 6.2 languages that I considered when developing my analysis and implementing it into the customization system. I then presented an evaluation of my system on data from five held-out languages 6.3. I showed that my library is able to handle a wide range of NI expression, but could be improved by increasing user customization in terms of promoted and doubled argument optionality. I also identified new phrase structure rules where the NCORP-MOD feature needs to be constrained. Finally, I showed that my library produces minimal overgeneration,

and identified areas where TDL editing can further constrain unwanted parses.

## Chapter 7

# CONCLUSION

This thesis has presented the work done to create a library for modeling noun incorporation in the Grammar Matrix customization system. I surveyed the literature on NI, drawing insight from previous analyses within and apart from the framework of HPSG. I then presented my own analysis of NI, accounting for crosslinguistic variation. I described how I implemented this analysis in the customization system and then evaluated its performance on three sets of natural and constructed languages.

### **7.1 Overview**

My analysis for NI accounts for instances where the IN fills the semantic role that an independent overt object would fill. I analyzed what I described as a three-way division in types of NI: Reduction, Promotion, and Modifier. Each type requires a pair of lexical rules, the first of which is shared across all three types. This lexical rule contributes the semantic relation for the IN. The secondary lexical rules each manipulate the first element on the verb's COMPS list, either making it empty or obligatorily unexpressed (Reduction NI), not identifying the INDEX between mother and daughter to allow a different noun to be represented there (Promotion NI), or identifying the INDEX and LTOP between mother and daughter to expose the values of these features to external modifiers. I also defined two unary phrase structure rules for stranded modifiers and doubled nouns which access the INDEX and LTOP of the IN to model the semantics of adnominal modifiers and hypernym/hyponym relationships respectively when combined in the syntax.

In implementation, I enabled the user to define incorporating stem lexical rule types within a verbal position class, and select which types of NI exist in the language they are

modeling from a series of options. Behind the scenes, my library maps the first part of the NI analysis to the user-defined PC, and generates a ‘ghost’ PC with lexical rule types for each type of NI indicated to be present in the language. This single position class approach takes the burden of defining two PCs for NI off of the user, but does introduce one known issue. For languages where one type of NI appears with one PC (such as transitive agreement) and another type of NI appears with a different one (such as intransitive agreement), it is possible for the incorrect intermediate types to be assigned in the grammar, resulting in overgeneration for sentences without NI. I documented this issue here, and provide further guidance in documentation to help users who encounter the same problem.

While creating this library, I interacted significantly with other existing libraries. I reported on some successful and some unsuccessful interactions, particularly with the valence change library (Curtis, 2018). In some languages, passivization and incorporation occur in the same construction. I was unable to effectively handle this interaction in the Inuktitut [iku] grammar due to the fact that the valence change library does not support passive operations on intransitive targets, and that I could not access the module for demoted subjects independently of the module for promoted objects. I was, however, able to model the interaction in the grammars for Southern Tiwa [tix] and Yaqui [yaq]. In the later case, this was made possible due to the lack of an overt subject in the passive, NI construction. This allowed me to use the subject removal operation from the valence change library.

## **7.2 Future Work**

I see many areas where future work on NI may occur. There are manifestations of NI known in the literature which I determined to be outside the scope of this initial project. These include the incorporation of patient-like subjects of intransitive verbs as well as of oblique arguments.

The evaluation of my system brought to light some shortcomings in the library as implemented. There were areas where I did not effectively constrain the NCORP-MOD feature which led to spurious ambiguity, particularly in phrase structure rules like that for posses-

sives and also on arguments impacted by valence-changing operations. I also found instances where giving more customization options to the user would have been better. Based on my review of the literature, I specified that promoted objects and doubled nouns should be overt, so these elements in their respective lexical rules are constrained to be [OPT –]. While this still seems like a good generalization, it would have been better to allow the user to say whether or not these elements can be dropped, and also to be able to specify more features on these elements such as their HEAD values. Another assumption I made is that, in a given language, Reduction NI will always either behave transitively or intransitively while I pursued a different analysis for modeling NI in Inuktitut, future work should investigate if it would be worth it to allow the user to enable both options in their grammar.

More work needs to be done to analyze the information structure impact of NI, perhaps including the *non-focus* relation found as a result of the transitive Reduction NI rule in some, or all, of the other NI types. More generally, an analysis for ditransitive verbs is needed which would then be able to extend the coverage for languages with a three-way agreement affix with the subject, IN, and promoted object in Promotion NI constructions. Computationally, it would be nice to pre-populate the orthography and predicate values of IS-LRIs from a class of nouns deemed to be able to incorporate. As it stands, any overt noun can appear as the object in doubling. Future work should constrain which doubled nouns can occur with which INs based on hyponym/hypernym relationships.

### 7.3 Summary

In closing, I have contributed a library which handles multiple types of noun incorporation. While there are areas to fix, improve, and expand, the system which I created performs well across multiple metrics for multiple languages. The analyses presented here will be a good foundation for expanding coverage of NI in the Grammar Matrix customization system in the future.

## BIBLIOGRAPHY

- Aikhenvald, A. Y. (2007). Typological distinctions in word-formation. *Language Typology and Syntactic Description*, 3, 1–65.
- Allen, B. J., Gardiner, D. B., & Frantz, D. G. (1984). Noun incorporation in Southern Tiwa. *International Journal of American Linguistics*, 50(3), 292–311.
- Baker, M. C. (1988). *Incorporation: A Theory of Grammatical Function Changing*. University of Chicago Press.
- Baker, M. C. (1996). *The Polysynthesis Parameter*. Oxford University Press.
- Baker, M. C., Aranovich, R., & Golluscio, L. (2005). Two types of syntactic noun incorporation: Noun incorporation in Mapudungun and its typological implications. *Language*, 81(1), 138–176.
- Ball, D. (2005). Tongan noun incorporation: Lexical sharing or argument inheritance. In S. Müller (Ed.), *Proceedings of the 12th International Conference on Head-Driven Phrase Structure Grammar, Department of Informatics, University of Lisbon* (pp. 7–27). CSLI Publications. <https://doi.org/10.21248/hpsg.2005.1>
- Bender, E. M., Drellishak, S., Fokkens, A., Poulson, L., & Saleem, S. (2010). Grammar customization. *Research on Language & Computation*, 8(1), 23–72. <http://dx.doi.org/10.1007/s11168-010-9070-1>
- Bender, E. M., Flickinger, D., & Oepen, S. (2002). The grammar matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. In J. Carroll, N. Oostdijk, & R. Sutcliffe (Eds.), *Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th international Conference on Computational Linguistics* (pp. 8–14).

- Bender, E. M., Poulson, L., Drellishak, S., & Evans, C. (2007). Validation and regression testing for a cross-linguistic grammar resource. In T. Baldwin, M. Dras, J. Hockenmaier, T. H. King, & G. van Noord (Eds.), *ACL 2007 Workshop on Deep Linguistic Processing* (pp. 136–143). Association for Computational Linguistics. <https://aclanthology.org/W07-1218/>
- Bogoras, W. (1917). *Koryak texts* (Vol. 5). EJ Brill.
- Bresnan, J., & Mchombo, S. A. (1995). The lexical integrity principle: Evidence from Bantu. *Natural Language & Linguistic Theory*, 13(2), 181–254.
- Bricker, V. R. (1978). Antipassive constructions in Yucatec Maya. *Papers in Mayan Linguistics*, 3–23.
- Chafe, W. L. (1976). *The Caddoan, Iroquoian and Siouan Languages* (Reprint 2013., Vol. 3). Mouton.
- Chung, S. (1978). *Case Marking & Grammatical Relations in Polynesian*. University of Texas Press. <http://catdir.loc.gov/catdir/enhancements/fy1502/78056993-b.html>
- Chung, S., & Ladusaw, W. A. (2004). *Restriction and Saturation*. MIT press.
- Churchward, C. M. (1953). *Tongan Grammar*. Oxford University Press.
- Copestake, A. (2002a). Definitions of typed feature structures. In S. Oepen, D. Flickinger, J. Tsujii, & H. Uszkoreit (Eds.), *Collaborative Language Engineering* (pp. 227–230). CSLI Publications.
- Copestake, A. (2002b). *Implementing Typed Feature Structure Grammars* (Vol. 110). CSLI publications Stanford.
- Copestake, A., Flickinger, D., Pollard, C., & Sag, I. A. (2005). Minimal recursion semantics: An introduction. *Research on Language and Computation*, 3(2-3), 281–332.
- Copestake, A., Lascarides, A., & Flickinger, D. (2001). An algebra for semantic construction in constraint-based grammars. *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, 140–147. <https://doi.org/10.3115/1073012.1073031>
- Crysmann, B., & Packard, W. (2012). Towards efficient hpsg generation for German, a non-configurational language. *Proceedings of COLING 2012*, 695–710.

- Curtis, C. M. (2018). *A Parametric Implementation of Valence-changing Morphology in the LinGO Grammar Matrix* [Master's thesis, University of Washington]. <http://hdl.handle.net/1773/41814>
- Davis, A. R., & Koenig, J.-P. (2000). Linking as constraints on word classes in a hierarchical lexicon. *Language*, 76(1), 56–91.
- Di Sciullo, A.-M., & Williams, E. (1987). *On the Definition of Word*. MIT Press.
- Dixon, R. M. W. (1988). *A Grammar of Boumaa Fijian*. University of Chicago Press.
- Drellishak, S. (2009). *Widespread but Not Universal: Improving the Typological Coverage of the Grammar Matrix* [Doctoral dissertation, University of Washington]. University of Washington. [http://gateway.proquest.com/openurl?url\\_ver=Z39.88-2004&rft\\_val\\_fmt=info:ofi/fmt:kev:mtx:dissertation&res\\_dat=xri:pqm&rft\\_dat=xri:pqdiss:3370489](http://gateway.proquest.com/openurl?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&res_dat=xri:pqm&rft_dat=xri:pqdiss:3370489)
- Drellishak, S., & Bender, E. M. (2005). A coordination module for a crosslinguistic grammar resource. In S. Müller (Ed.), *The Proceedings of the 12th International Conference on Head-Driven Phrase Structure Grammar, Department of Informatics, University of Lisbon* (pp. 108–128). CSLI Publications. <http://csli-publications.stanford.edu/HPSG/6/>
- Facundes, S. d. S. (2000). *The Language of the Apurinã People of Brazil (Maipure/Arawak)* [Doctoral dissertation, State University of New York at Buffalo].
- Flickinger, D. (2000). On building a more efficient grammar by exploiting types. *Natural Language Engineering*, 6(1), 15–28.
- Flickinger, D. (2011). Accuracy vs. robustness in grammar engineering. In E. M. Bender & J. E. Arnold (Eds.), *Language From a Cognitive Perspective: Grammar, Usage, and Processing* (pp. 31–50, Vol. 201). CSLI Publications Stanford.
- Flickinger, D., Bender, E. M., & Oepen, S. (2003). *MRS in the LinGO Grammar Matrix: A Practical User's Guide* [ms].
- Fortescue, M. D. (1984). *West Greenlandic*. Croom Helm London.
- Friesen, D. (2017). *A Grammar of Moloko*. Language Science Press.

- Gerdtts, D. B. (2017). Incorporation. In A. Spencer & A. M. Zwicky (Eds.), *The Handbook of Morphology* (pp. 84–100). John Wiley & Sons, Ltd. <https://doi.org/https://doi.org/10.1002/9781405166348.ch4>
- Goodman, M. W. (2013). Generation of machine-readable morphological rules from human-readable input. *University of Washinton Working Papers in Linguistics*, 30. [http://depts.washington.edu/uwwpl/vol30/goodman\\_2013.pdf](http://depts.washington.edu/uwwpl/vol30/goodman_2013.pdf)
- Haugen, J. D. (2008). Noun incorporation in Uto-Aztcán. In *Morphology at the Interfaces* (pp. 117–162, Vol. 117). John Benjamins Publishing Company.
- Jelinek, E. (2014). Voice and transitivity as functional projections in Yaqui. In A. Carnie & H. Harley (Eds.), *Pronouns, Presuppositions, and Hierarchies: The Work of Eloise Jelinek in Context* (1st ed., pp. 331–360). Routledge.
- Koenig, J.-P., & Davis, A. R. (2001). Sublexical modality and the structure of lexical semantic representations. *Linguistics and Philosophy*, 24(1), 71–124.
- Lee, J. (1987). *Tiwi Today: A Study of Language Change in a Contact Situation*. Dept. of Linguistics, Research School of Pacific Studies, Australian National University, 1987.
- Levinson, S. C. (2018). Introduction: Demonstratives: Patterns in diversity. In S. C. Levinson, S. Cutfield, M. J. Dunn, N. J. Enfield, & S. Meira (Eds.), *Demonstratives in Cross-linguistic Perspective* (pp. 1–42). Cambridge University Press.
- Malouf, R. (1999). West Greenlandic noun incorporation in a monohierarchical theory of grammar. In A. K. Gert Webelhuth & J.-P. Koenig (Eds.), *Lexical and Constructional Aspects of Linguistic Explanation* (pp. 47–62). CSLI Publications, Stanford.
- Massam, D. (2001). Pseudo noun incorporation in Niuean. *Natural Language & Linguistic Theory*, 19(1), 153–197.
- Massam, D. (2009). Noun incorporation: Essentials and extensions. *Language and Linguistics Compass*, 3(4), 1076–1096.
- McKay, G. R. (1975). *Rembarrnga: A language of Central Arnhem Land*. The Australian National University (Australia).

- Merlan, F. C. (1976). Noun incorporation and discourse reference in Modern Nahuatl. *International Journal of American Linguistics*, 42, 177–191. <https://api.semanticscholar.org/CorpusID:144771505>
- Mithun, M. (1984). The evolution of noun incorporation. *Language*, 60(4), 847–894. <http://www.jstor.org/stable/413800>
- Müller, S., Abeillé, A., Borsley, R. D., & Koenig, J.-P. (Eds.). (2024). *Head-Driven Phrase Structure Grammar: The handbook, Second revised edition*. Language Science Press. <https://doi.org/10.5281/zenodo.13637708>
- Nielsen, E. K. (2018). *Modeling Adnominal Possession in the LinGO Grammar Matrix* [Master's thesis, University of Washington]. <http://hdl.handle.net/1773/42433>
- Oates, L. F. (1964). A tentative description of the Gunwinggu language (of Western Arnhem Land). *Oceania Linguistic Monographs*.
- O'Hara, K. (2008). *A Morphotactic Infrastructure for a Grammar Customization System* [Master's thesis, University of Washington]. University of Washington.
- Osborne, C. R. (1974). *The Tiwi Language: Grammar, Myths and Dictionary of the Tiwi Language Spoken on Melville and Bathurst Islands, Northern Australia*. Australian Institute of Aboriginal Studies.
- Pacchiarotti, S., & Zúñiga, F. (2025). Noun incorporation in Bribri. *International Journal of American Linguistics*, 91(4), 477–507.
- Parks, D. R. (1977). *Caddoan Texts* (1st ed.). The University of Chicago Press.
- Polinskaja, M. S., & Nedjalkov, V. P. (1987). Contrasting the absolutive in Chuckchee: Syntax, semantics and pragmatics. *Lingua*, 71(1-4), 239–269.
- Polinsky, M. (1990). Subject Incorporation: Evidence from Chukchee. In K. Dziwirek, P. Farrell, & E. Mejias-Bikandi (Eds.), *Grammatical relations: A cross-theoretical perspective* (pp. 349–364).
- Pollard, C., & Sag, I. A. (1994). *Head-Driven Phrase Structure Grammar*. University of Chicago Press.

- Rosen, S. T. (1989). Two types of noun incorporation: A lexical analysis. *Language*, 65(2), 294–317. <http://www.jstor.org/stable/415334>
- Runner, J. T., & Aranovich, R. (2003). Noun incorporation and rule interaction in the lexicon. In S. Müller (Ed.), *Proceedings of the 10th International Conference on Head-Driven Phrase Structure Grammar, Michigan State University* (pp. 359–379). CSLI Publications. <https://doi.org/10.21248/hpsg.2003.20>
- Sadock, J. M. (1985). Autolexical syntax: A proposal for the treatment of noun incorporation and similar phenomena. *Natural Language & Linguistic Theory*, 3(4), 379–439.
- Salas, A. (1992). El Mapuche o Araucano: Fonología, gramática y antología de cuentos. *Editorial MAPFRE*.
- Saleem, S. (2010). *Argument optionality: A New Library for the Grammar Matrix Customization System* [Master's thesis, University of Washington]. University of Washington.
- Saleem, S., & Bender, E. M. (2010). Argument optionality in the LinGO grammar matrix. In C.-R. Huang & D. Jurafsky (Eds.), *Coling 2010: Posters* (pp. 1068–1076). Coling 2010 Organizing Committee. <https://aclanthology.org/C10-2123/>
- Smeets, I. (2008). *A Grammar of Mapuche*. De Gruyter Mouton. <https://doi.org/doi:10.1515/9783110211795>
- Song, S. (2017). *Modeling information structure in a cross-linguistic perspective*. Language Science Press.
- Spencer, A. (1995). Incorporation in Chukchi. *Language*, 439–489.
- Trimble, T. (2014). *Adjectives in the LinGO Grammar Matrix* [Master's thesis, University of Washington]. <http://hdl.handle.net/1773/27512>
- Wasow, T. (1977). Transformations and the lexicon. *Formal Syntax*, 327–360.
- Wasow, T. (1980). Major and minor rules in lexical grammar. *Lexical Grammar*, 285, 321.
- Weir, H. (1990). Incorporation in Nadëb. *Amazonian Linguistics: Studies in Lowland South American Languages*, 321–363.
- Yuan, M. (2025). Morphological conditions on movement-chain resolution: Inuktitut noun incorporation revisited. *Natural Language & Linguistic Theory*, 43(1), 509–562.

Zamaraeva, O., Curtis, C., Emerson, G., Fokkens, A., Goodman, M., Howell, K., Trimble, T., & Bender, E. M. (2022). 20 years of the grammar matrix: Cross-linguistic hypothesis testing of increasingly complex interactions. *Journal of Language Modelling*, *10*(1), 49–137. <https://doi.org/10.15398/jlm.v10i1.292>