A cross-linguistic analysis of noun incorporation in HPSG

Emily Luedke
University of Washington

1 Introduction

Noun Incorporation (NI) is the morphological attestation of a noun on a verb stem where the noun bears some semantic relation to the verb (such as direct object). In this paper, I present an analysis of three structurally different types of direct object NI (henceforth NI), accounting for sub-variation within each type. This analysis presents a typology of NI types — rather than a typology of languages with NI. To test my analysis, I implemented the lexical and phrase structure rules described below as a library for the LinGO Grammar Matrix customization system (Bender, Flickinger & Oepen 2002; Bender et al. 2010; Zamaraeva et al. 2022), and tested how well it performed over test suites using grammars created for five held-out natural languages.

2 Background

2.1 NI literature

In her influential typological survey, Mithun 1984 describes four functionally different types of noun incorporation. Type I is simple compounding, where a noun is incorporated into a verb to create a word for a common activity. In Type II, a noun is incorporated to foreground another argument which syntactically takes the place that the incorporated noun (IN) would have had in an unincorporated construction. Type III involves the incorporation of a noun which has previously been introduced in the discourse. Finally, Type IV NI serves to qualify an external noun phrase, where the IN and external noun are co-referential.

Other literature on NI has centered on a distinction between languages with Compounding NI or Classifier NI (Rosen 1989). Compounding NI is seen to decrease the transitivity of a verb, whereas Classifier NI does not. Two characteristics of Classifier NI languages are doubling and stranding. Stranding is a construction in which some modifier of the IN appears 'stranded' outside of the verb, and doubling refers to instances where the IN is co-referential with an identical or more specific external noun.

2.2 HPSG NI analyses

Previous work on NI within the framework of HPSG (Pollard & Sag 1994) include three language-specific analyses (Malouf 1999, Ball 2005, Koenig & Michelson 2021) and an analysis of Compounding and Classifier NI (Runner & Aranovich 2003).

Malouf 1999 analyzes denominal verbs — a related but distinct phenomenon from NI — in West Greenlandic (Kalaalisut) [kal] with a mixed-category analysis in which the resulting construction shares characteristics of both the verbal and nominal input.

Ball 2005 extends the work of Malouf 1999 with an Argument Inheritance Analysis for NI in Tongan [ton]. A lexical rule creates a verb from a noun and verb input with the selectional constraints of both the nominal and verbal elements.

Koenig & Michelson 2021 provide an analysis for incorporation in Oneida [one] where the concatenation of a noun root and a verb stem results in a verb stem with the INDEX of the verbal element, and where the INDEX of the nominal root has a meaningful relation to the verb stem daughter.

Runner & Aranovich 2003 define separate type constraints for Compounding and Classifier NI constructions to account for the lack of an overt object argument in the syntax in both cases but differences in morphological behavior. The authors do not present an analysis of doubling or stranding, which are fundamental elements of Classifier NI languages elsewhere in the literature (c.f. Baker 1996; Baker, Aranovich & Golluscio 2005; Rosen 1989).

2.3 The LinGO Grammar Matrix

I created this analysis in the context of building a library for the Grammar Matrix customization system. The Grammar Matrix is a project which supports users in building implemented grammars using Head-Driven Phrase Structure Grammar (HPSG) (Pollard & Sag 1994) and Minimal Recursion Semantics (MRS) (Copestake et al. 2005). The customization system contains stored analyses of different linguistic phenomena that can be customized for a language to produce a grammar fragment written in the TDL formalism. These grammars can be compiled using DELPH-IN software for parsing and generation tasks.

3 Data

Instead of providing a language-based typological analysis of NI along the lines of Runner & Aranovich 2003,

I provide a construction level analysis along the lines of Mithun's typology (Mithun 1984). I define and break my analysis into three types: Reduction NI, Promotion NI, and Modifier NI. Reduction NI refers to constructions in which a noun is incorporated and there is no other overt expression of this argument. Promotion NI refers to constructions in which a noun is incorporated, and some other element is promoted to the syntactic position of the verb's direct object. Modifier NI refers to constructions in which some external element modifies the IN. This type accounts for the occurrence of both doubling and stranding.

In Reduction NI constructions, the overt transitivity of the sentence is reduced such that the verb does not combine syntactically with a direct object. (1) and (2) illustrate this type of NI.[1]

(1) ətləg-ən qaa-nmə-gʔe
father-ABS reindeer(INC)-kill-AOR.3SG
'The father killed a reindeer.' [ckt] (Polinsky 1990 via Gerdts 1998:88)

(2) bi-seuan-mu-ban
1sS/BO-man-see-PAST
'I saw men' [tix] (Allen, Gardiner & Frantz 1984 via Baker, Aranovich & Golluscio 2005:141)

These two examples differ in the behavior of the rest of the clause. Wherein (1) the clause displays morphology typical of intransitive constructions in Chukchi [ckt] (i.e. absolutive marking on and verbal agreement only with the subject), (2) patterns like a typical Southern Tiwa [tix] transitive verb (i.e. verbal agreement with both the subject and IN).

Promotion NI constructions do not result in a change in transitivity of the clause because another argument is promoted to object position. (3) illustrates the promotion of the possessor of the IN in Mapudungun [arn], and (4) illustrates this position being filled by an oblique, in this case benefactive, argument in Chukchi [ckt].

(3) Juan ngilla-waka-fi-y Pedro
Juan buy-cow-3O-IND.3sS Pedro
'Juan bought Pedro's cow.' [arn] (Baker, Aranovich & Golluscio 2005:167)

(4) tumg-e ekək kaynə-nmə-nen
friend-ERG son-(ABS) bear-kill-AOR.3SG.3SG
'The friend killed the bear for his son.' [ckt] (Polinsky 1990 via Gerdts 1998:91)

In Modifier NI, some external element modifies the IN. (5) shows an example of doubling in Mohawk [moh], where the IN *itsy* 'fish' is accompanied by the more specific noun *rabahbót* 'bullhead.' In the Southern Tiwa [tix] example in (6), the IN *diru* 'chicken' is modified by the stranded demonstrative *yede* 'that.'

(5) sha'teku ni-kuti rabahbót
eight PART-ZPLS bullhead
wa-hv-[i]tsy-a-hninu-' ki rake-'niha
FACT-MSGS-fish-∅-buy-PUNC this my-father
'My father (fish-)bought eight bullheads.' [moh] (Baker 1996:310 via Baker, Aranovich & Golluscio 2005:149)

(6) yede a-diru-k'ar-hi
that 2sS/AO-chicken-eat-FUT
'You will eat that chicken.' [tix] (Allen, Gardiner & Frantz 1984:296 via Baker, Aranovich & Golluscio 2005:150)

4 Analysis

I analyze the process of noun incorporation as a pair of verbal lexical rules. The first lexical rule is shared across all types of NI. An example of this rule is sketched in Figure 1.[2] This is an inflecting lexical rule which appends the orthographic form of the IN as defined by a user of the Grammar Matrix customization system to the verb's STEM feature. The rule contributes the noun-relation and its associated quant-relation to the semantics. The rule also identifies the ARG0 of the noun-relation with the INDEX value of the first element on the verb's COMPS list, connecting the INDEX of the IN to a semantic role in the verb's EP in the MRS.

The second lexical rule is specific to the type of NI, and modifies the COMPS list of the verb to account for the presence or absence of promoted, stranded, or doubled material. Each of these are non-inflecting rules, so the STEM feature of the verb remains unchanged. In implementation, I make use of the Morphotactics library (O'Hara 2008; Goodman 2013) which contains functionality for modeling dependencies between

[1] A more detailed analysis of the data, including unincorporated paraphrases and worked examples of the lexical rules presented, can be found in my master's thesis (Luedke 2026).

[2] In this feature structure as well as others, I abbreviate some paths for the sake of space and readability. For example, "SYNSEM ... COMPS" refers to the path "SYNSEM|LOCAL|CAT|VAL|COMPS."

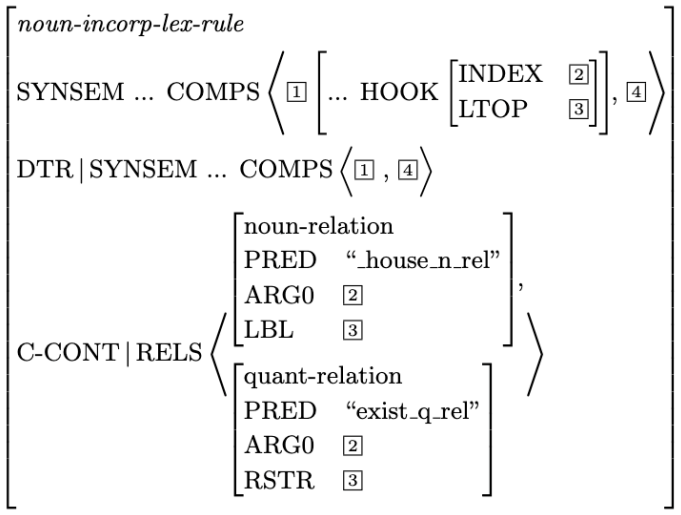

Figure 1: Common NI lexical rule

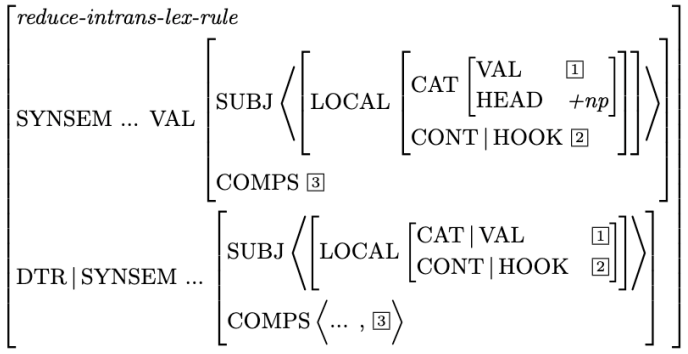

Figure 2: Intransitive Reduction NI lexical rule

lexical rules. This ensures that both the general and type-specific lexical rules apply together and in the correct order.

4.1 Reduction NI lexical rules

The lexical rule for Reduction NI in which the clause behaves intransitively is sketched in Figure 2. This rule only identifies the value of VAL between the mother's and daughter's SUBJ list instead of the entire SUBJ list in order to account for differences between transitive and intransitive subjects in a language (e.g. case). Copying up VAL ensures that the verb continues to select for a saturated argument. The HEAD value on the mother is constrained to be *+np* which is a disjunctive type used in the Grammar Matrix for nouns and adpositions. The rule identifies the mother's COMPS list with the daughter's COMPS|REST. If there was only one element on the daughter's COMPS list, then that on the mother will be empty.

The lexical rule for transitive Reduction NI sketched in Figure 3 constrains the first element on the verb's

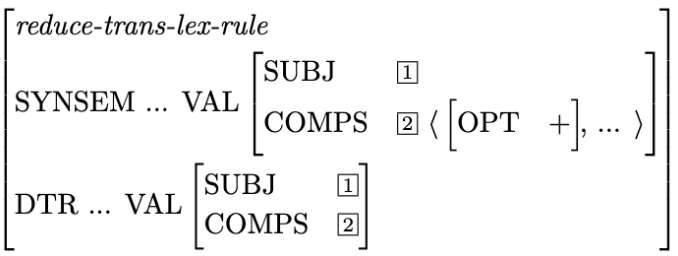

Figure 3: Transitive Reduction NI lexical rule

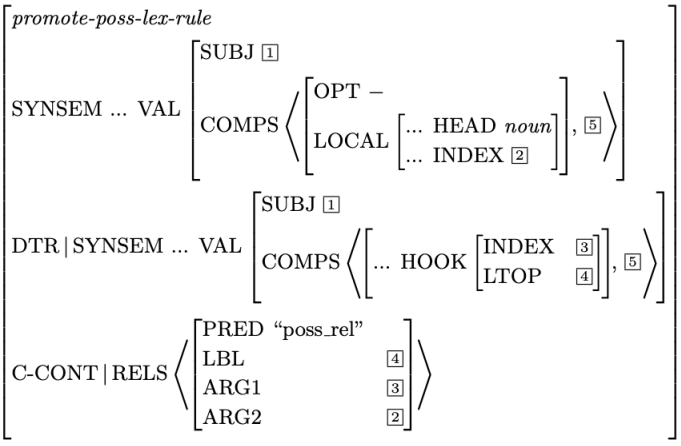

Figure 4: Promotion NI lexical rule: promoted possessor

COMPS list to be [OPT +].[3] This feature prevents the verb from combining with an object in the syntax, but allows features associated with the IN to still be accessible on the COMPS list (e.g. person and number features).

4.2 Promotion NI lexical rules

The lexical rule for NI constructions with promoted possessors is sketched in Figure 4. This rule does not identify the INDEX value of the first element on the verb's COMPS list between mother and daughter, allowing the verb to pick up a new object. To obtain the correct possession semantics, this rule introduces a poss_rel relation which identifies its ARG1 with the INDEX of the IN, and its LBL with the IN's LTOP. The ARG2 of this relation is identified with the INDEX of the promoted object.

The Promotion NI rule for promoted oblique arguments is sketched in Figure 5. As above, this rule does not identify the INDEX of the first element on the verb's COMPS list between mother and daughter, allowing the verb to pick up a new argument in the syntax. This

[3] OPT is a feature created for the Matrix Argument Optionality library (Saleem & Bender 2010).

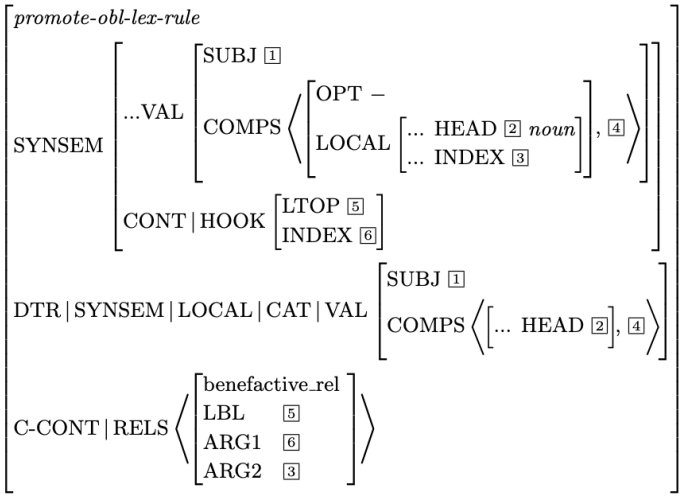

Figure 5: Promotion NI lexical rule: promoted oblique argument

rule adds a relation to the semantics which accounts for oblique arguments. The LBL and ARG1 of the relation is shared with the LTOP and INDEX of the verb, respectively. The ARG2 of this relation is identified with the INDEX of the promoted oblique argument.[4]

4.3 Modifier NI lexical and phrase structure rules

My analysis of Modifier NI involves both a lexical rule for the verb which has gone through the general NI lexical rule as well as a phrase structure rule that applies to the external element modifying the IN. In order to constrain what may modify an IN and to allow for the verb to select for an appropriate modifying argument, I define a new boolean feature NCORP-MOD (short for noun incorporation modifier) which is a feature inside of CAT.

The lexical rule for doubling is sketched in Figure 6. This rule applies to the verb and specifies that the NCORP-MOD value on the first element of the mother's COMPS list is positive. The rule also identifies the LTOP and INDEX values of the first element on the verb's COMPS list (i.e. the LTOP and INDEX of the IN) between mother and daughter. This identification is necessary to expose these values when the [NCORP-MOD +] element is picked up by the *head-comp* rule in the syntax.

The phrase structure rule for doubled nouns is sketched in Figure 7. This rule applies to noun phrases which modify an IN, and specifies that the NCORP-MOD

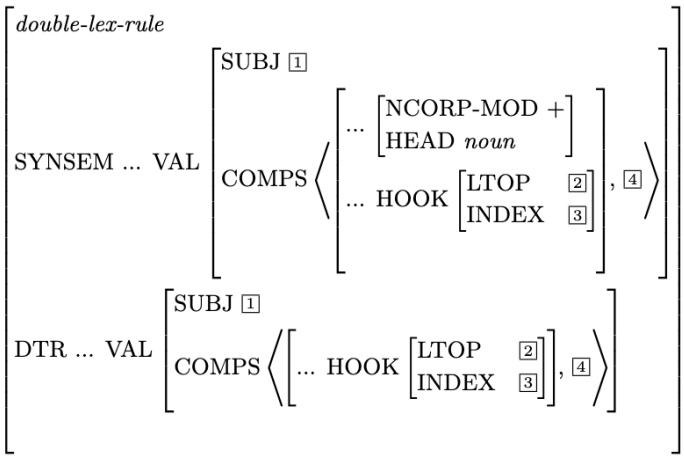

Figure 6: Modifier NI lexical rule: doubling

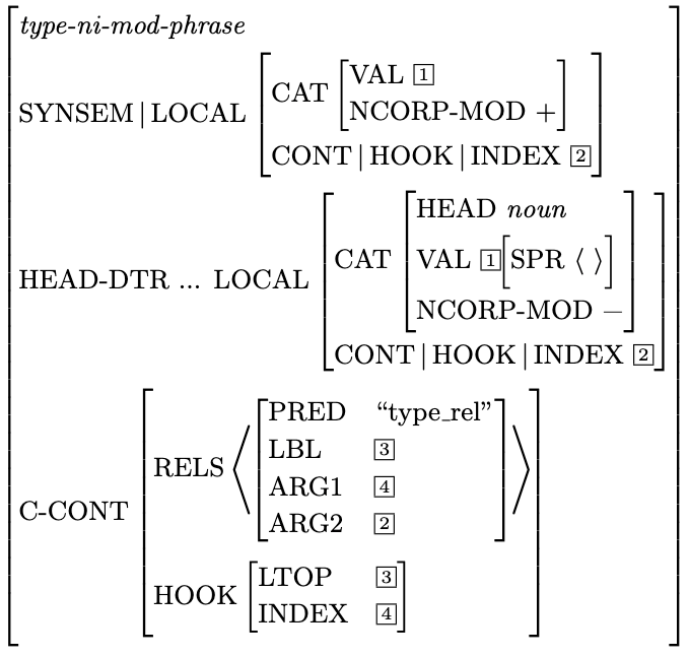

Figure 7: Doubling phrase structure rule

[4] At the time of writing, ditransitive verbs are not yet supported in the Grammar Matrix customization system. Future work on noun incorporation will involve accounting for verbs which already have more than one arguments on their COMPS list.

value is negative on the daughter but positive on the mother. The rule contributes a relation to the semantics with a PRED value of "type_rel" which refers to the hypernym/hyponym relationship of INs and doubled nouns. This relation's LBL and ARG1 are identified with the LTOP and INDEX of the IN, accessed via the C-CONT|HOOK feature. The ARG2 value is identified with the INDEX of the doubled noun. In the MRS, the ARG1 and ARG2 of this relation map to the ARG0s of EPs for the incorporated and doubled nouns respectively.[5]

The lexical rule for stranding is sketched in Figure 8. This rule also specifies that the NCORP-MOD value on the first element of the mother's COMPS list needs to be positive. Additionally, this element is constrained to have a HEAD value of *+jp* which is a disjunctive type used in the Grammar Matrix that includes both adjectives and adopositions. The LTOP and INDEX values of this element are again identified between mother and daughter to expose these features to achieve the correct semantics when combining with the stranded modifier in the syntax.

The phrase structure rule for stranded modifiers is sketched in Figure 9. This rule applies to adjectives and specifies that the NCORP-MOD value on the mother of the rule is positive. This rule identifies the INDEX of the IN — accessed via the C-CONT.HOOK feature — with the INDEX of the element on the adjective's MOD list. It also identifies the LTOP of the IN with that of the adjective, which is consistent with the semantics of adjectives elsewhere in Matrix grammars (Trimble 2014).

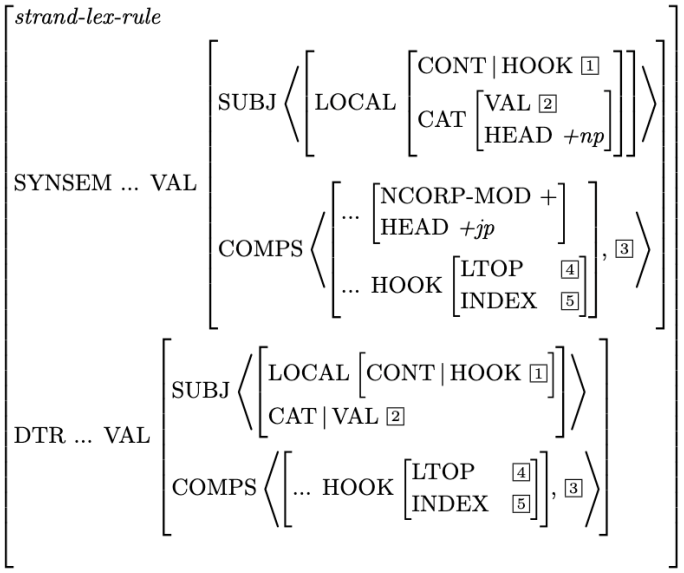

Figure 8: Modifier NI lexical rule: stranding

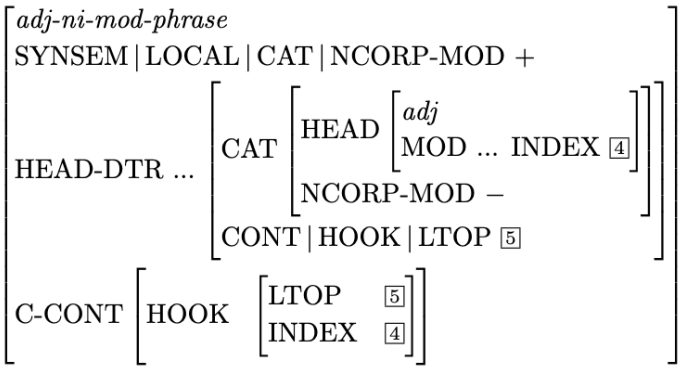

Figure 9: Stranding phrase structure rule

5 Implementation and evaluation

I implemented the analysis presented above as a library for the Grammar Matrix customization system. I tested the analysis on five held-out languages, measuring the coverage, overgeneration, and parsing ambiguity for positive and negative strings in test suites created for each language. The results for each grammar are shown in Table 1.

6 Conclusion

I present an analysis of three different types of NI, including instances of variation within these types. This work expands on that of Runner & Aranovich 2003 to provide a novel analysis for both stranding and doubling. Languages can select from this typology of NI types to include the relevant rules in an implemented grammar. Evaluation of the implementation of this analysis shows that it extends well to unseen data, with %100 coverage for 4/5 grammars. Future work can extend this analysis to account for less common types of NI such as subject and oblique argument incorporation.

[5]This rule currently does not rule out pragmatically unexpected sentences where the IN is not related to the doubled noun. Future work can extend this rule by creating a hierarchy of entities such that the rule constrains the sentences to be pragmatically sound.

References

Allen, Barbara J., Donna B. Gardiner & Donald G. Frantz. 1984. Noun incorporation in Southern Tiwa. *International Journal of American Linguistics* 50(3). 292–311.

Baker, Mark C. 1996. *The Polysynthesis Parameter*. Oxford University Press.

Baker, Mark C, Roberto Aranovich & Lucia Golluscio. 2005. Two types of syntactic noun incorporation:

Language	Cov.	Overgen.	Amb.
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 1: Held-out language grammar performance

noun incorporation in Mapudungun and its typological implications. *Language* 81(1). 138–176.

Ball, Douglas. 2005. Tongan noun incorporation: Lexical sharing or argument inheritance. In Stefan Müller (ed.), *Proceedings of the 12th International Conference on Head-Driven Phrase Structure Grammar, Department of Informatics, University of Lisbon*, 7–27. Stanford, CA: CSLI Publications. https://doi.org/10.21248/hpsg.2005.1.

Bender, Emily M., Scott Drellishak, Antske Fokkens, Laurie Poulson & Safiyyah Saleem. 2010. Grammar customization. *Research on Language & Computation* 8(1). 23–72. https://doi.org/10.1007/s11168-010-9070-1.

Bender, Emily M., Dan Flickinger & Stephan Oepen. 2002. The grammar matrix: an open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. In John Carroll, Nelleke Oostdijk & Richard Sutcliffe (eds.), *Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th international Conference on Computational Linguistics*, 8–14. Taipei, Taiwan.

Copestake, Ann, Dan Flickinger, Carl Pollard & Ivan A. Sag. 2005. Minimal recursion semantics: an introduction. *Research on Language and Computation* 3(2-3). 281–332.

Gerdts, Donna B. 1998. Incorporation. In Andrew Spencer & Arnold M. Zwicky (eds.), *The Handbook of Morphology*, chap. 4, 84–100. Basil Blackwell.

Goodman, Michael Wayne. 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.

Koenig, Jean-Pierre & Karin Michelson. 2021. Specialized-domain grammars and the architecture of grammars: Possession in Oneida. *Journal of linguistics* 57. 635–674.

Luedke, Emily. 2026. *Next, Incorporate the Flour: a Recipe for Modeling Noun Incorporation in the LinGO Grammar Matrix*. University of Washington MA thesis.

Malouf, Robert. 1999. West Greenlandic noun incorporation in a monohierarchical theory of grammar. In Andreas Kathol Gert Webelhuth & Jean-Pierre Koenig (eds.), *Lexical and Constructional Aspects of Linguistic Explanation*, 47–62. CSLI Publications, Stanford.

Mithun, Marianne. 1984. The evolution of noun incorporation. *Language* 60(4). 847–894. http://www.jstor.org/stable/413800.

O'Hara, Kelly. 2008. *A Morphotactic Infrastructure for a Grammar Customization System*. University of Washington MA thesis.

Polinsky, Maria. 1990. Subject Incorporation: Evidence from Chukchee. In Katarzyna Dziwirek, Patrick Farrell & Errapel Mejias-Bikandi (eds.), *Grammatical relations: a cross-theoretical perspective*, 349–364.

Pollard, Carl & Ivan A Sag. 1994. *Head-Driven Phrase Structure Grammar*. University of Chicago Press.

Rosen, Sara Thomas. 1989. Two types of noun incorporation: a lexical analysis. *Language* 65(2). 294–317. http://www.jstor.org/stable/415334.

Runner, Jeffrey T. & Raul Aranovich. 2003. Noun incorporation and rule interaction in the lexicon. In Stefan Müller (ed.), *Proceedings of the 10th International Conference on Head-Driven Phrase Structure Grammar, Michigan State University*, 359–379. Stanford, CA: CSLI Publications. https://doi.org/10.21248/hpsg.2003.20.

Saleem, Safiyyah & Emily M. Bender. 2010. Argument optionality in the LinGO grammar matrix. In Chu-Ren Huang & Dan Jurafsky (eds.), *Coling 2010: posters*, 1068–1076. Beijing, China: Coling 2010 Organizing Committee. https://aclanthology.org/C10-2123/.

Trimble, Thomas. 2014. *Adjectives in the LinGO Grammar Matrix*. Seattle: University of Washington MA thesis. http://hdl.handle.net/1773/27512.

Zamaraeva, Olga, Chris Curtis, Guy Emerson, Antske Fokkens, Michael Goodman, Kristen Howell, T.J. Trimble & Emily M. Bender. 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.