Focus case outside of Austronesian: An analysis of Kolyma Yukaghir

Olga Zamaraeva
University of Washington

Emily M. Bender
University of Washington

Proceedings of the 21st International Conference on
Head-Driven Phrase Structure Grammar

University at Buffalo

Stefan Müller (Editor)

2014

Stanford, CA: CSLI Publications

pages 176–196

Zamaraeva, Olga & Emily M. Bender. 2014. Focus case outside of Austronesian: An analysis of Kolyma Yukaghir. In Stefan Müller (ed.), *Proceedings of the 21st International Conference on Head-Driven Phrase Structure Grammar, University at Buffalo*, 176–196. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2014.10.

Abstract

Case is traditionally approached as a lexical phenomenon in HPSG. The LinGO Grammar Matrix customization system, an HPSG-based grammar engineering toolkit and also a typological meta-resource, includes several options for case assignment, and one of them, 'focus case', assumes that case of the participants in basic clauses is handled via lexical rules rather than lexical entries. This phenomenon was previously only attributed to a group of Austronesian languages, and thus the focus case differed from all other case options in the Matrix which were attested for across language families. Our analysis of Kolyma Yukaghir, a nearly extinct language of North-Eastern Russia, shows that focus case can be successfully used outside of Austronesian family and therefore that the option is more universal than it was previously thought.

1 Introduction

Since at least Pollard & Sag, 1994, case assignment in HPSG has been seen as a lexical phenomenon: "assignment of case to complements [...] is simply treated as part of subcategorization" (*Ibid.* p.30). Simple subcategorization isn't enough to capture all facts of case assignment (Przepiórkowski, 1996), but it still serves as a useful core: On the one hand, lexicalist analyses of valence alternations add lexical rules (e.g. for passive or causative alternations) which produce new case assignments as part of new subcategorization frames (e.g. Müller, 2001) and on the other hand, many authors have proposed a distinction between lexical and structural case (e.g. Heinz & Matiasek, 1994; Przepiórkowski, 1996), allowing the actual morphological form of structural cases to be sensitive to the syntactic environment. However, even with all of these extensions, the analysis of case still crucially involves the subcategorization frames of verbs.

Drellishak, 2009 adds a library for case to the LinGO Grammar Matrix customization system (Bender et al., 2002, 2010).[1] This library, based on a rigorous review of typological literature, provides nine choices of general case system (including nominative-accusative, ergative-absolutive, several kinds of splits, and none) which in turn provide basic case subcategorization frames appropriate to the language type as well as facilities for defining additional case values and additional case frames (e.g. quirky case). The ninth of these types, called 'focus-case' is of particular interest here: Unlike the other case systems, it does not involve underlying case frames for verbs. Rather, verbs are required to undergo inflectional rules which constrain the case on the arguments. It also stands out from the others, as far as Drellishak was able to determine, in only being attested in one language family (Austronesian).

In this paper, we describe an analysis of the case system of Kolyma Yukaghir (ISO 639-3: yux), a language of North-Eastern Russia. Strikingly, though this

[1] Drellishak's library only concerns lexical case.

system has some clear differences to that found in Tagalog and other Austronesian languages, the focus case option of Drellishak's library supports an elegant implementation. In §2, we outline the fundamentals of Drellishak's analysis. §3 presents a brief sketch of Yukaghir. We present our analysis (implemented in a Grammar Matrix based grammar) in §4 and evaluate it against a hand-constructed testsuite in §5. Finally, §6 describes how the grammar can be extended to handle the interaction of case marking and definiteness.

2 Focus case in the Grammar Matrix

In Austronesian languages (including Tagalog), the case frame of the verb depends on an inflectional marker that picks out which argument is in 'focus'.[2] This is illustrated in (1)–(2), where the AGENT.FOCUS and PATIENT.FOCUS markers on the verb indicate which role the *ang*-marked element should be interpreted as filling.

(1) Bumili ang babae ng baro
bought-AGENT.FOCUS FOCUS woman PATIENT dress
'The woman bought a dress.' [tgl] (Drellishak, 2009, p.54)

(2) Bimili ng babae ang baro
bought-PATIENT.FOCUS AGENT woman FOCUS dress
'A/the woman bought the dress.' [tgl] (Drellishak, 2009, p.54)

Drellishak's analysis of this system leaves the case of arguments underspecified in the lexical entries for the verbs, and then uses lexical rules to fill in case frames. More specifically, these lexical rules form a required 'position class' that every verb must go through.[3] These lexical rules attach the affixes for AGENT.FOCUS, PATIENT.FOCUS etc. while also constraining the case values of all elements of the verb's arguments structure appropriately.

3 Kolyma Yukaghir

Yukaghir languages[4] are considered either a small isolated family or a distant relative of Uralic languages (Fortescue, 1996, p.17). Kolyma Yukaghir is nearly extinct (estimates vary from just 5 to 300-400 speakers, depending on the definition of fluency) (Maslova, 2003, p.1). It is basically a SOV agglutinating language (Fortescue, 1996, p.17), though systematic deviations from SOV word order are attested

[2] It is not clear whether this actually corresponds to the information-structural notion of focus in these languages, as a 'focused' constituent other than the verb is required in every clause; see §3.2.

[3] A position class in a Grammar Matrix-derived grammar is a type describing a set of lexical rules which take the same inputs and in turn can serve as inputs to the same set of further lexical rules (Goodman, 2013).

[4] There are two varieties: Kolyma and Tundra Yukaghir. They are typologically very similar, but whenever we say "Yukaghir" in this paper, we mean Kolyma Yukaghir.

(Maslova, 2003, p.341) and pragmatics often determines the word order as well, so that even in simple clauses, different word orders are possible (Maslova, 2003, p.17). It has singular and plural number, first, second, and third person, no gender system, and fairly developed morphology. There is inflection for future and non-future tense, and a periphrastic construction for past tense, using a nominalized verb form (*Ibid*. p.6). There are also aspect distinctions (including imperfective, ingressive, resultative, and habitual), and four major moods (inferential, prospective, irrealis, and periphrastic prospective) (*Ibid*. p.6). Subjects and objects are easily dropped when recoverable from context (*Ibid*. p.9), unless they are in focus (*Ibid.* p.326).

Focus, which is expressed grammatically, is one of the most notable features of the language. Of particular interest is the role of information structure in syntax of the clause. In literature about Yukaghir, the well-established and traditional use of the term 'focus' is justified primarily by "grammaticalized association of Focus role with canonical 'focus-presupposition' contexts, where the information about the situation being described is (directly or indirectly) 'activated' by the time of utterance or can be viewed as a part of extra-linguistic context, the referential identity of one participant being the only unknown piece of information about the situation" (Maslova, 2005, p.600). Thus in Yukaghir, the term 'focus' seems to basically correspond to the usual information structure sense, as summarized, for example, in Song, 2014. The focus marking on the nouns is referred to as predicative case in Maslova, 2003.

3.1 Case in Kolyma Yukaghir

Yukaghir exhibits a fairly complex system of case marking. Maslova distinguishes 9 cases for nouns and 12 for pronouns, of which the following seem to be most involved in the basic intransitive and transitive verb patterns: nominative, predicative, accusative, instrumental, and a form called 'pronominal accusative' which we analyze as object non-focus marker ('NFO') for non-3rd person pronouns (in contrast to nouns which use the unmarked form both in positions requiring nominative case and where pronouns would take NFO). As we will describe further below, the choice of the case frame in simple clauses typically depends not on the particular verb, but rather on the information structure of the clause, as well as on the person value of the subject in transitive clauses.

3.1.1 Intransitive clauses

In intransitive clauses, the subject can either be in focus or it can be neutral to focus. If the subject is focused, it takes predicative case, and the verb takes the marker *-l* glossed SF for 'subject focus' (3). When an intransitive subject is not focused, it takes nominative case and the verb agrees with it in person and number (4).[5]

[5]Abbreviations used in Yukaghir examples:

(3) tāt touke-lek jede-l
CA dog-PRED appear-SF
'Then {a dog}$_{Foc}$ appeared.' [yux] (Maslova, 2003, p.9)

(4) met ejre-je
I.NOM walk-1SG
'I walked.' [yux]

3.1.2 Transitive clauses

In transitive clauses, the verb registers whether or not the object is in focus (OF) (5). The marking of this information interacts with the marking for subject agreement: There are are two sets of subject agreement markers, one used when the object is focused and one used when it is not. The choice of the case frame for most verbs that we considered depends on whether the verb has the OF marker, whether the subject is 3rd person (6) or not (7), and additionally whether the object is definite (8). The paradigms are summarized in Table 1. The subject is always in nominative case; in focused transitive clauses, the object is always in predicative case. In non-focused transitive clauses, if the subject is 1st or 2nd person, the object appears in the NFO (non-focused object) form: the so-called 'pronominal accusative' for first and second person pronouns and the zero-marked form (equivalent to nominative) for nouns and third person pronouns (9). If the subject is 3rd person, a definite object will be accusative (e.g. (8)) and an indefinite instrumental (e.g. (6)).

(5) tet-ek aŋči-nu-ŋile
you-PRED search-IPFV-3PL.OF

12PER	1st and 2nd Person
1SG	1st Person Singular
ACC	Accusative case
CA	connective adverbial
FUT	Future
NONFUT	Nonfuture
INSTR	Instrumental case
INTR	Intransitive
IPFV	Imperfective aspect
NEG	Negation
NF	Non-focus
NFO	Non-focused Object case; 'Pronominal Accusative' in (Maslova, 2003)
NOM	Nominative case
OF	Object Focus
PERNUM	Person and Number marking
PL	Plural
PRED	Predicative (Focus) case
SF	Subject Focus
TR	Transitive
NON3PL	Not 3rd plural

All glosses except 'NFO' are from Maslova, 2003. Examples without citations are constructed by the authors and have been verified by a Yukaghir expert.

OF		Non-Focus	
S 3rd	S non-3rd	S 3rd	S non-3rd
NOM-PRED		NOM-ACC/INSTR	NOM-NOM/NFO

Table 1: Transitive clause case frames

'It is {you}$_{Foc}$ whom they are seeking.' [yux] (Maslova, 2003, p.153)

(6) tudel tolow-le kudde-m
he.NOM deer-INSTR kill-3SG
'He killed a deer.' [yux] (Maslova, 2003, p.10)

(7) met tolow kudede
I.NOM deer.NOM kill.1SG
'I killed a deer.' [yux] (Maslova, 2003, p.10)

(8) tudel met kønme-gele juø-m
he.NOM my friend-ACC see-3SG
'He saw my friend.' [yux] (Maslova, 2003, p.10)

(9) met tet-ul juø
I.NOM you-NFO see.1SG
'I saw you.' [yux] (Maslova, 2003, p.10)

We leave ditransitive, chained, and non-finite clauses out of the discussion. Generally only finite clauses can be marked in this way for focus. There are other case frames which involve other cases such as ablative, as well as classes of verbs which require locative, but the most basic intransitive and transitive patterns can be summarized as above.

3.2 Focus in Kolyma Yukaghir and in Tagalog

Not surprisingly, given the great distance separating their geographic distributions, Kolyma Yukaghir and Tagalog are quite different typologically and in particular in the details of their focus case systems. Tagalog uses adpositional marking to mark nouns for case (Comrie, 1989, p.121); Yukaghir uses inflectional marking. The Yukaghir verbal markers implicated in this system also carry person and number information, where the Tagalog markers do not (Cruz & Shkarban, 1966). Finally, where every Tagalog clause must have an element which is marked with the so-called 'focus marker' *ang* (Comrie, 1989, p.121), Yukaghir, allows for clauses where no argument bears focus marking. We believe that this means that the Yukaghir system is more likely to actually be marking information structural contrasts: It seems implausible that a language would systematically disallow verbs to be in (narrow) focus, but if Tagalog *ang* were to strictly represent information structural focus, narrow verbal focus would be impossible. Yukaghir, on the other hand, allows clauses where the NPs are not marked for focus, leaving open the possibility of the verb carrying focus.

In the initial development of our Yukaghir grammar, the focus case option was passed over, at least in part because it was presented as only appropriate for Austronesian languages. We ended up developing by hand a system that was quite close to Drellishak's analysis, and so went back to the customization system and created a grammar fragment using the focus case option. This is the grammar fragment presented in §4 and evaluated in §5 below, and it shows that despite its typological and geographical distance from Austronesian, when viewed through the framework of HPSG, Yukaghir shares a certain typological similarity with Tagalog and its kin.

4 Analysis of Kolyma Yukaghir in the Grammar Matrix

In order to test our analysis of the case system, we needed our grammar fragment to handle enough other basic facts to parse our test sentences which exemplify choice of case frame in intransitive and transitive clauses. We created the fragment through the Grammar Matrix customization system (Bender et al., 2002, 2010).[6] In the process, we made a few simplifying decisions: In particular, we said that word order is 'free', though in some cases some orders are not possible.[7] We also did not implement any of the moods and only implemented habitual aspect as an example. Generally, in this fragment we only try to cover the basic case assignment in simple clauses. For number, person and gender the Matrix provides suitable options (sg/pl, 1/2/3 and no gender). It is also possible to model sentential negation (which is simple negation expressed by prefix *el-* on the verb), but this doesn't affect the analysis of case.

4.1 Case

We picked the focus case option for Kolyma Yukaghir despite the note in the Grammar Matrix customization system questionnaire which presented it as relevant only to the Austronesian family. Unlike the other case system options in the Grammar Matrix customization system, this one does not provide a set of default argument structures with case frames pre-defined. Instead, it supports the implementation of verbal lexical rules which fill in case requirements on the verb's arguments. As described below, this option supports an effective analysis of the Yukaghir system.

We restricted our analysis to the subset of cases described in the previous section (nominative, predicative, accusative, instrumental, and NFO). The customization system questionnaire requires specification of 'focus case', 'A-case' and 'O-case' when the focus case option is chosen.[8] We specified these as predicative,

[6]The grammar fragment and our testsuite are available online at http://depts.washington.edu/uwcl/matrix/yukaghir/.

[7]Maslova (2003) says that the language is basically SOV, though generally other orders are possible. In our testsuite, only VSO and VOS were deemed ungrammatical by the Yukaghir expert who gave the judgments.

[8]A and O refer to the subject/object of a transitive verb, respectively.

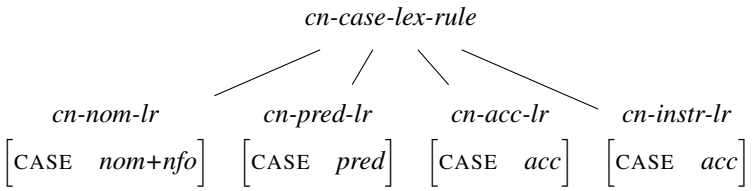

Figure 1: Noun Case lexical rules hierarchy

accusative and nominative respectively. Instrumental and NFO were listed as 'additional cases'. Later in the morphology section, we were able to model the case frames based on this choice. The customized grammar was able to handle the facts of Yukaghir case (as detailed in §5 below) without further modification, with one exception: The customization system does not yet provide facilities for constraining discourse/cognitive status (Borthen & Haugereid, 2005) of arguments, and so our grammar overgenerates with respect to the distribution of accusative and instrumental objects. This can of course be remedied by hand-editing of the grammar (see §6).

4.2 Lexicon

We populated the lexicon with all the personal pronouns and a few basic common nouns, to be able to test example sentences. We did not include possessives or other parts of speech. This means that for testing, we used constructed examples such as 'He saw a friend' rather than 'He saw my friend'. In order to model a definite NP, we included a demonstrative determiner, so it is possible to parse a sentence 'He saw this friend'. We included a few basic verbs: *ejre* ('walk'), *jede* ('appear'), *juø* ('see'), *kudede* ('kill'). Intransitive and transitive verbs are the only verb classes that we included in the grammar fragment (these are the only types supported at present by the Grammar Matrix customization system questionnaire); a more extensive grammar would of course require more classes (for example, verbs of cognition, qualitative verbs).

4.3 Morphology

The main goal of our grammar fragment is to implement case requirements on verbs and the associated marking of case on nouns (and pronouns), which requires lexical rules for both. We assume a morphological analyzer to account for some of the morphophonological phenomena, which are described in Maslova (2003). Accordingly, our grammar targets morpheme-segmented, regularized forms such as those shown in the examples in this paper.

4.3.1 Noun and Pronoun inflection rules

The key observation for the analysis for nouns and pronouns is that the distribution of the NFO case marker on the pronouns is exactly like the zero nominative case marker for nouns as direct objects (zero marker on the object noun if the subject is non-3rd person). This is exemplified in (7) and (9), repeated here as (10) and (11). Compare with (12) and (13) to see that the surface forms of the NFO and the NOM are the same for nouns, but different for first and second person pronouns. Examples (14), (15), and (16) show the 3rd person pronoun forms (which are the same for NOM, NFO as well as PRED).

(10) met tolow kudede
I.NOM deer.NOM kill.1SG
'I killed a deer.' [yux] (Maslova, 2003, p.10)

(11) met tet-ul juø
I.NOM you-NFO see.1SG
'I saw you.' [yux] (Maslova, 2003, p.10)

(12) tolow met-gele kudede-m
deer.NOM I.ACC kill-3SG
'The deer killed me.' [yux]

(13) tet met-ul juø-mek
you.NOM I.NFO see-2SG
'You killed me.' [yux]

(14) tudel met-gele kudede-m
he.NOM I.ACC kill-3SG
'He killed me.' [yux]

(15) met tudel-gele kudede
I.NOM he.ACC kill.1SG
'I killed him.' [yux]

(16) met tudel kudede-me
I.NOM he.NOM kill-OF.1SG
'I killed him$_{Foc}$.' [yux]

Thus, it makes sense to say the zero-marker (Maslova's 'nominative') marks nouns for nominative ('A-case' in the focus case library terms) as well as for NFO (see Figure 1). In contrast, pronouns must have separate lexical rules for nominative and NFO (Figure 2). Then, in the verb lexical rules section, it is sufficient to constrain the *tr-nf* object to be in NFO case (shown in Figure 4).

We analyze both the (surface) instrumental and accusative as marking nouns as [CASE *acc*].[9] This allows for the instrumental forms but does not properly

[9]This analysis leaves open the possibility of a separate, homophonous, instrumental case rule that produces [CASE *instr*] nouns with definiteness unspecified.

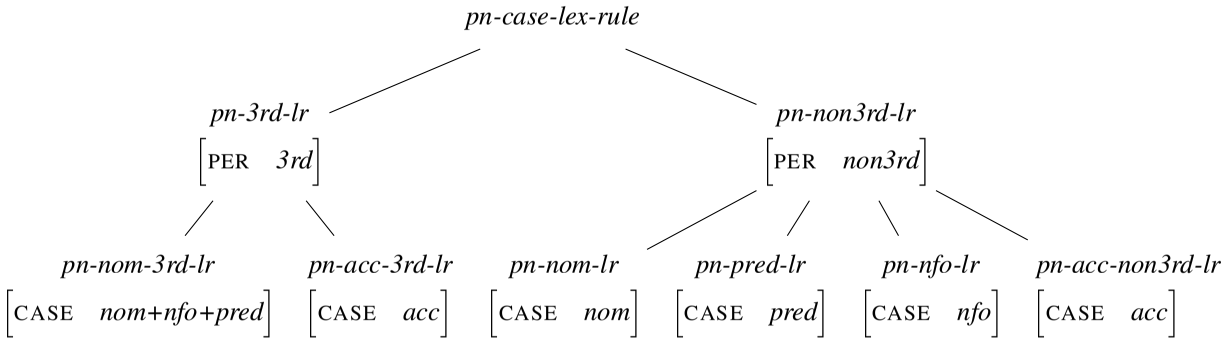

Figure 2: Pronoun Case lexical rules hierarchy

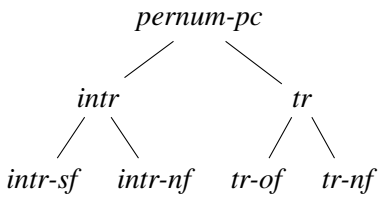

Figure 3: Verb lexical rules

limit their distribution: The additional constraints required (on definiteness) can be added directly via tdl editing, but are not at present supported by the Grammar Matrix customization system (see section 6).

4.3.2 Verb inflection rules

Unlike other options for case systems in the Grammar Matrix customization system, focus case does not provide a set of argument structure values with case specified to use in the definition of verb classes. Rather, picking the option creates case frames that can be used in verb lexical rules as constraint on the verb's argument structure. (The high level hierarchy for these rules is presented in Figure 3; further detail is given below.) In case of Yukaghir, the automatically generated 'nom-pred' case frame option can be used in modeling the *tr-of* lexical rule type. In general, three lexical rule types give rise to the majority of the lexical rules: intransitive non-focused subject (*intr-nf*), transitive focused object (*tr-of*), and transitive non-focused object (*tr-nf*). Since there is only one marker for any verb with a focused subject, intransitive focused subject (*intr-sf*) type can be realized via a single lexical rule, much like Drellishak's treatment of Tagalog transitive clauses (Drellishak, 2009, p.66). In order to model the various patterns in transitive clauses found in Yukaghir, additional lexical rules are required, as discussed below.

In the basic morphology that we consider the position classes are chained as follows (Maslova, 2003, p.149):

(17) NEG-stem-ASPECT-TENSE-AGR

However, in verbs agreeing with 3PL subjects, the tense marker is in between two markers which both mark the verb for person and number.[10] Therefore it is more practical to assume an additional position class, which comes before tense, classifies the verb's subject as 3rd person plural or not, and is typically still accompanied by a person and number marker after the tense marker:

(18) juø-ŋi-te-m
see-3PL-FUT-3PL

[10]We differ from Maslova in analyzing these as both marking both person and number (i.e. an instance of multiple exponence), as this leads to a correct association of form and morphosemantic features in our system.

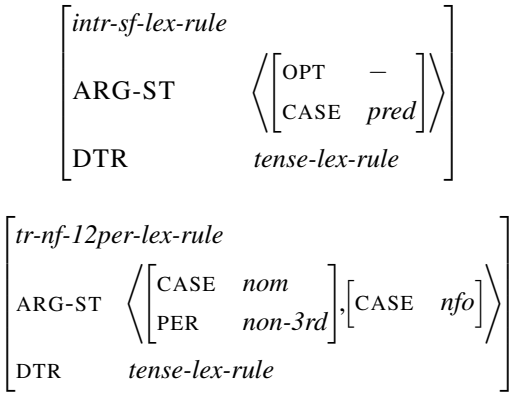

Figure 4: Sample lexical rules

'They will see.' [yux] (Maslova, 2003, p.140)[11]

(19) juø-0-te-m
see-NON3PL-FUT-3SG
'(S)he will see.' [yux] (Maslova, 2003, p.140)

(20) ejre-ŋi-0-0
walk-3PL-NONFUT-3PL
'They walk.' [yux] (Maslova, 2003, p.140)

(21) ejre-ŋi-te-j
walk-3PL-FUT-3PL
'They will walk.' [yux] (Maslova, 2003, p.140)

With the 3PL position class in place, and with the future/non-future position class taking the output of the 3PL rule as input, we are ready to create the rules involved in determining case frames. We do this in terms of a hierarchy where the supertypes *intr-sf*, *intr-nf*, *tr-of*, and *tr-nf* constrain the case frames. *Intr-sf* is instantiated by just one lexical rule instance. The others all have subtypes describing full paradigms of person/number values on the subject, and thus all four supertypes correspond to the AGR position class in (17).

Implementing the paradigm for intransitives is relatively simple: There is one rule for subject focus (with no agreement distinctions), which attaches the subject-focus marker to the verb and constrains its subject to be [CASE *pred*], and a set of rules for non-focused subjects that indicate agreement in person and number and constrain the subject to be [CASE *nom*].

One source of complexity in the transitive paradigms is that the 3rd person marker depends also on whether the tense is future or not, since if it is, the above mentioned 3PL position class marker is nonzero while the AGR marker shows less variety.

[11]Examples (18)–(21) come from a morphological table which we used for reference and were not originally in the form of IGT.

The transitive branch is further complicated by the split on the subject's person value in terms of the case frame. In addition, the *intr-sf* rule also specifies that the subject is [OPT −] (cannot be dropped), and the *tr-of* rule does the same for its object. Figure 4 illustrates how the case frames are constrained for *intr-sf* and *tr-nf-12per*, respectively. Figures 5–7 show the type hiearchies and key constraints contributed by each type for these lexical rules.

The choices file implementing this analysis includes 699 individual choices (pieces of information). Among these, many are dedicated to defining the 35 lexical rules that handle case assignment and person/number agreement with the subject (32 in the AGR position class and 3 in the 3PL position class). While it may seem that this is a lot, it is in fact a manageable analysis of four distinct paradigms offered in Maslova's verb morphology table (Maslova, 2003, p.140). Furthermore, we note that this results in a working, testable grammar fragment for Yukaghir.

4.3.3 Information structure and argument optionality

The Grammar Matrix customization system also provides a library for modeling information structural constraints (Song, 2014). As noted above, it appears that the focus-related morphology in the verb does mark focus in the information structural sense. Using Song's library we are able to model this by selecting the option that creates affixes as focus markers and then adding the specification that the *intr-sf* and *tr-of* rules constrain their subject and object, respectively, to be in focus. This has the effect that the rules add an element to the verb's ICONS list representing the information structural meaning (in Song's terminology) of focus on the relevant argument. More specifically, giving this specification to the customization system causes the type *tr-of*, supertype to all of the object focus rules, to inherit from *add-icons-comp-rule*,[12] shown in Figure 8 and then to further constrain the particular type of ICONS element added to be *focus*.

An added benefit of modeling the information structural meaning is that it interacts with another constraint provided by the information structure library to correctly model the generalization that the focused arguments (subject of SF verbs, object of OF verbs) cannot be dropped, despite fairly free argument optionality within Yukaghir. Song models this as a general constraint on focus cross-linguistically: Focused arguments must be overtly realized and so the argument drop rules (non-branching productions provided by the argument optionality library (Saleem & Bender, 2010) constrain the argument being discharged to be [ICONS-KEY *non-focus*]. Without any further specification on our part, the customized grammar thus correct rules out examples like (22).

(22) *jede-te-l
appear-FUT-SF
Intended: '(someone) will appear.' [yux]

[12]The type *add-icons-comp-rule* is provided in the Matrix core grammar as part of Song's information structure library.

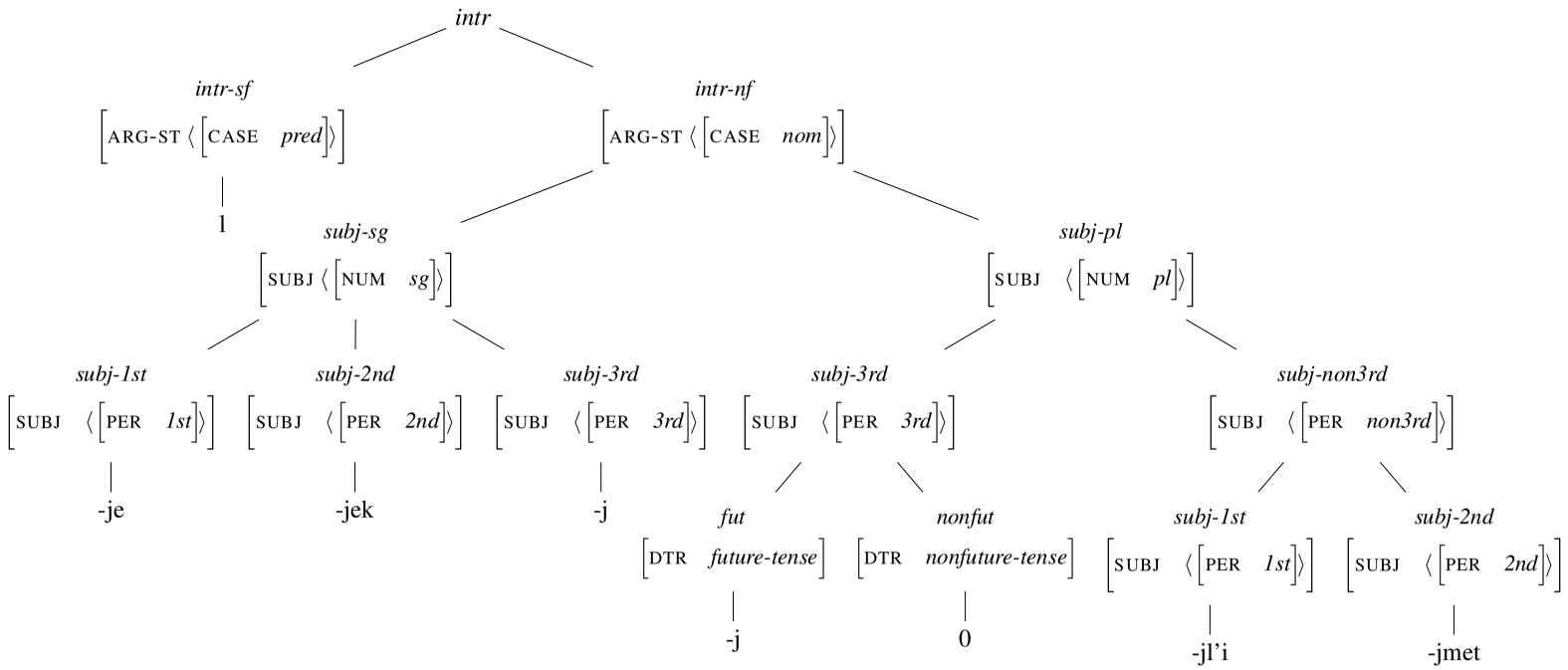

Figure 5: Intransitive verbs lexical rules

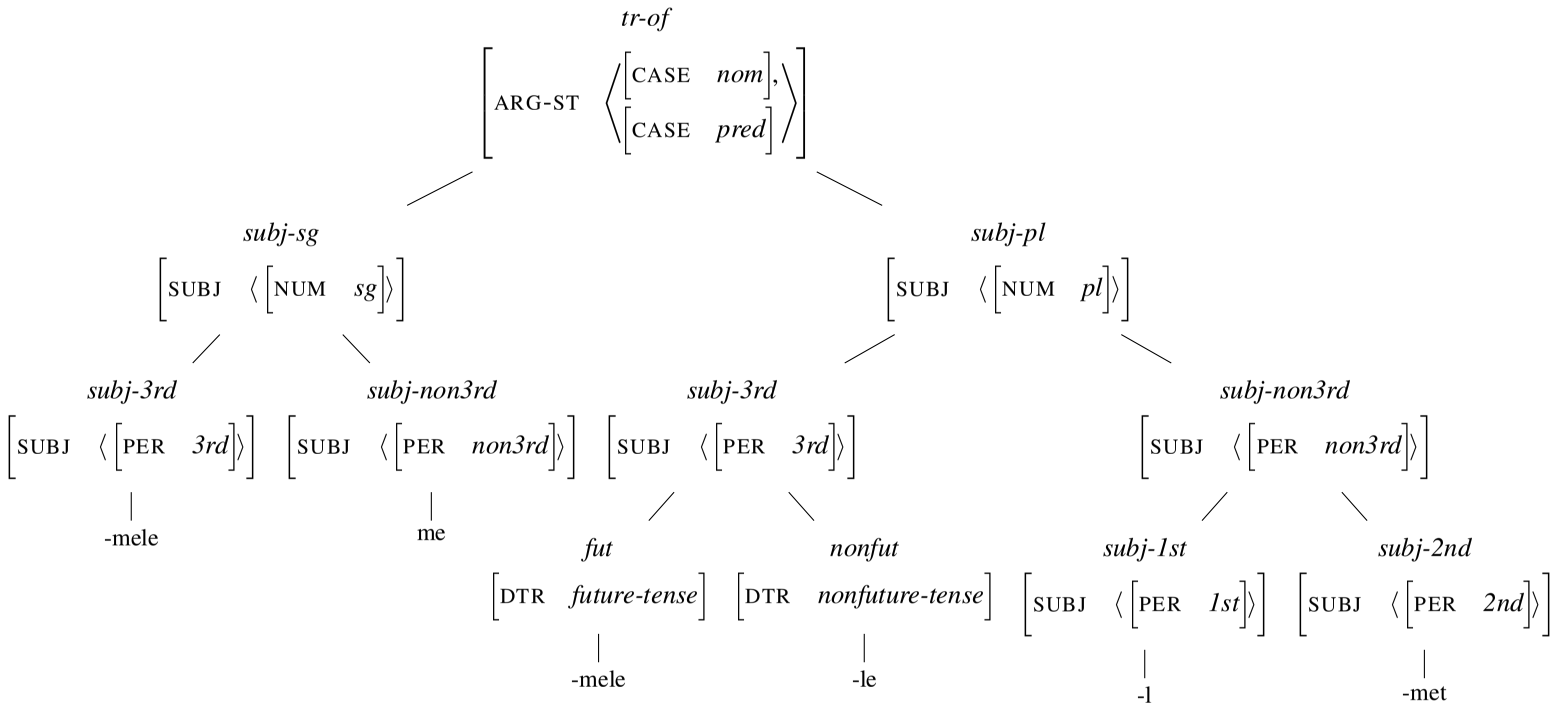

Figure 6: Transitive-object-focus verbs lexical rules

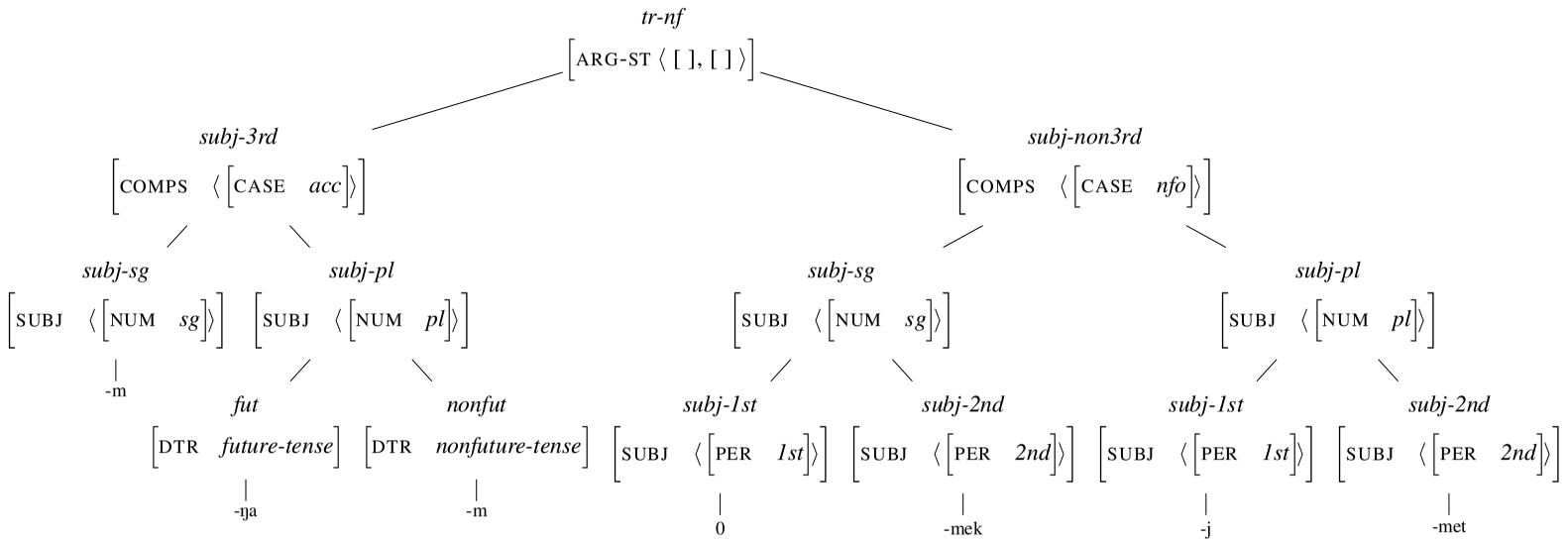

Figure 7: Transitive-object-non-focus verbs lexical rules

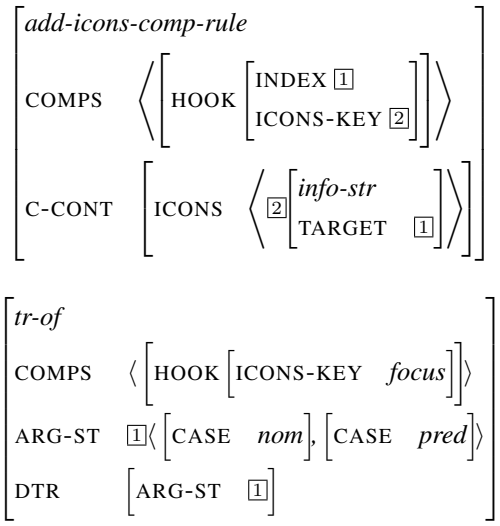

Figure 8: Constraints on ICONS modeling information structure of focused objects

5 Evaluation

We used 161 sentences,[13] 81 grammatical, 80 ungrammatical, to test the grammar that we obtained via the Matrix, using the LKB (Copestake, 2002) and [incr tsdb()] (Oepen, 2001). The performance of the grammar is summarized in Table 2.[14]

Coverage	Overgeneration
84.0 %	6.2%

Table 2: Yukaghir Grammar Performance

Inspection of the testsuite shows that the unanalyzed sentences (undergeneration) are the ones that represent phenomena beyond the scope of our grammar fragment (ditransitives, copula, attributive forms).

The overgeneration falls into two classes. On the one hand, our grammar allows VSO and VOS orders, marked as ungrammatical in the testsuite, due to our simplifying decision to use the free word order option. On the other hand, we also overgenerate because the grammar does not model the correlation between case and definiteness on non-focused objects. Specifically, as shown in (6) and (8) (repeated

[13] Most of the sentences are very simple, only involving a subject, and object, and a verb. Due to the lack of such basic examples in the available literature, most test examples are constructed from more complex sentences by removing possessives (and changing accusative case on the object to instrumental where appropriate), adverbs, and other words not directly involved in the transitive or intransitive pattern. All sentences have been verified by a Yukaghir expert.

[14] These numbers reflect changes to the testsuite compared to the results presented at the conference. In particular, we removed some of the more complex sentences with larger vocabulary, since they were not supposed to be covered by the grammar, and changed the word order from SOV to Free, after consulting with our Yukaghir expert. The version of the grammar and testsuite as evaluated can be found at http://depts.washington.edu/uwcl/matrix/yukaghir/.

here as (23) and (24)), indefinite non-focused objects are marked with instrumental case, while definite non-focused objects bear accusative. Furthermore, as shown in (25), definite objects require an overt determiner.[15] Our grammar fragment, however, accepts (25) and furthermore does not reflect the definiteness contrast in the semantics.

(23) tudel tolow-le kudde-m
he.NOM deer-INSTR kill-3SG
'He killed a deer.' [yux] (Maslova, 2003, p.10)

(24) tudel met kønme-gele juø-m
he.NOM my friend-ACC see-3SG
'He saw my friend.' [yux] (Maslova, 2003, p.10)

(25) *tudel tolow-gele kudde-m
he.NOM deer-ACC kill-3SG
Intended: 'He killed a/the deer.' [yux]

6 Extension: Definiteness of the Accusative

The grammar fragment we created with the customization system does not handle the definiteness distinction discussed immediately above because the Grammar Matrix customization system does not yet provide a library for definiteness marking. However, the Grammar Matrix's core grammar does provide support for adding such an analysis to the grammar by hand, based on Borthen & Haugereid's (2005) analysis of cognitive status of references (Bender & Goss-Grubbs, 2008).

To do this, we need to add demonstrative determiners, as in Figure 9. This type is instantiated by lexical entries such as *tiŋ* 'this', which contribute two elementary predications, an existential quantifier and the demonstrative (adjective) relation. Most relevantly here, they also constrain their SPEC's COG-ST (cognitive status) value to *activ+fam*, the value typical of definite NPs. Meanwhile, the lexical rules for instrumental and accusative case must also be enhanced. The former adds the information that the noun's index is [COG-ST *type-id*]. This value, typical of indefinite NPs, is incompatible with the constraints on the determiner. The latter constrains the noun's specifier requirement to [OPT −], requiring an overt determiner,[16] successfully blocking (25).

7 Conclusion

The case system of Yukaghir is relatively complex: the case frames of verbs depend on both the person of the subject (familiar from languages with split-ergativity)

[15] In this matter, opinions differ, and it is not clear whether examples like (25) would indeed be rejected by the speakers or not. We followed Maslova (2003) and assumed that (25) would be rejected.

[16] There are no indefinite determiners in the language; indefinites are bare NPs.

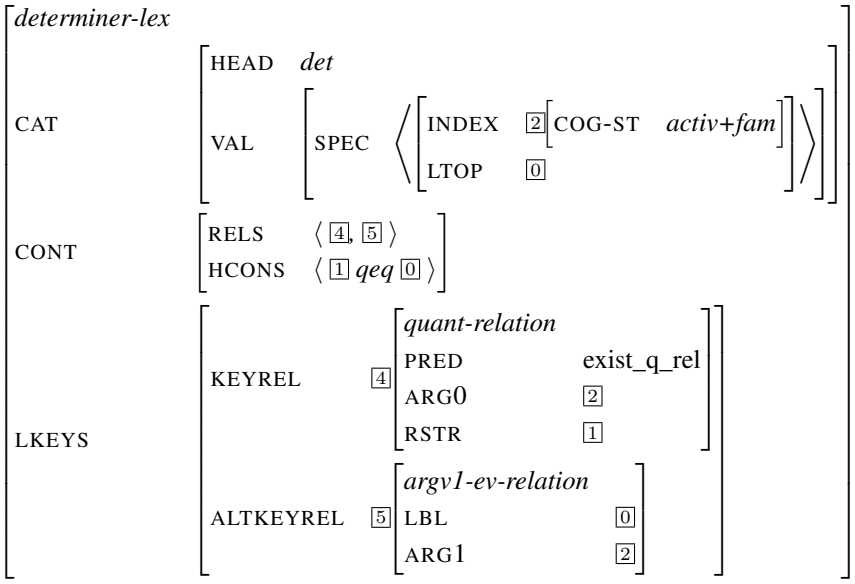

Figure 9: Constraints on *determiner-lex*

and on whether the verb bears focus marking. On the surface, the pattern does not immediately resemble that of the Austronesian languages which inspired the 'focus-case' option in Drellishak's case library for the Grammar Matrix. However, on closer inspection, Drellishak's analysis provides the core of an elegant account of this complex system, whether or not the Austronesian notion of 'focus' corresponds to the traditional information structure sense.

We have tested that analysis by creating a grammar fragment with the Grammar Matrix customization system. This fragment is able to handle all of the patterns described above, with the exception of the association between accusative case and definiteness, for which we offer a solution that requires some editing of the grammar by hand. This grammar fragment can be further extended as well: We find that the customization system's information structure library (Song, 2014) provides a suitable analysis for the information structural effects of focus marking in Yukaghir.

References

Bender, Emily M., Scott Drellishak, Antske Fokkens, Laurie Poulson & Safiyyah Saleem. 2010. Grammar customization. *Research on Language & Computation* 1–50. 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.

Bender, Emily M. & David Goss-Grubbs. 2008. Semantic representations of syntactically marked discourse status in crosslinguistic perspective. In *Proceedings of STEP 2008*, College Publications.

Borthen, Kaja & Petter Haugereid. 2005. Representing referential properties of nominals. *Research on Language and Computation* 3(2). 221–246.

Comrie, Bernard. 1989. *Language universals and linguistic typology: Syntax and morphology*. Chicago, IL: University of Chicago Press.

Copestake, Ann. 2002. *Implementing typed feature structure grammars*. Stanford, CA: CSLI Publications.

Cruz, Manuel & Lina I. Shkarban. 1966. *Tagal'slij yazyk*. Moscow, Russia: Nauka.

Drellishak, Scott. 2009. *Widespread but not universal: Improving the typological coverage of the Grammar Matrix*: University of Washington dissertation.

Fortescue, Michael. 1996. Grammaticalized focus in Yukaghir. is it really grammaticalized and is it really focus? In *Content, expression and structure: Studies in Danish structural grammar*, 17–38. John Benjamins Publishing.

Goodman, Michael Wayne. 2013. Generation of machine-readable morphological rules with human readable input. *UW Working Papers in Linguistics* 30.

Heinz, Wolfgang & Johannes Matiasek. 1994. Argument structure and case assignment in German. *German in Head-Driven Phrase Structure Grammar* 46. 199–236.

Maslova, Elena. 2003. *A grammar of Kolyma Yukaghir* Mouton Grammar Library (Book 27). Berlin: Mouton De Gruyter.

Maslova, Elena. 2005. Information structure in Tundra Yukaghir and typology of focus structures. In M. M. Jocelyn Fernandez-Vest (ed.), *Les langues ouraliennes aujourd'hui*, 599–610. Honore Champion.

Müller, Stefan. 2001. The passive as a lexical rule. In *Proceedings of the 7th international conference on HPSG*, 247–266.

Oepen, Stephan. 2001. [incr tsdb()] — Competence and performance laboratory. User manual. Technical report Saarbrücken, Germany.

Pollard, Carl & Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar* Studies in Contemporary Linguistics. Chicago, IL and Stanford, CA: The University of Chicago Press and CSLI Publications.

Przepiórkowski, Adam. 1996. Case assignment in Polish: Towards an HPSG analysis. In Claire Grover & Enric Vallduví (eds.), *Studies in HPSG*, vol. 12 Edinburgh Working Papers in Cognitive Science, 191–228. Centre for Cognitive Science, University of Edinburgh.

Saleem, Safiyyah & Emily M. Bender. 2010. Argument optionality in the lingo grammar matrix. In *Coling 2010: Posters*, 1068–1076. Beijing, China: Coling 2010 Organizing Committee. http://www.aclweb.org/anthology/C10-2123.

Song, Sanghoun. 2014. *A grammar library for information structure*: University of Washington dissertation.