Focus Case outside of Austronesian: An Analysis of Yukaghir

Olga Zamaraeva and Emily M. Bender
University of Washington

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 provides nine choices of general case system (including none, nominative-accusative, ergative-absolutive, and several kinds of splits) 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 (for e.g. quirky case).

2 Focus case in the Grammar Matrix

The ninth of these types, called 'focus-case' is of particular interest here: It was added be-cause the Austronesian-style case marking system could not be assimilated (on the HPSG analysis) to any of the other types. 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]

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

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

Drellishak's analysis of this system leaves the case of arguments underspecified in the lexical entries for the verbs, and then requires a position class[3] for lexical rules that each constrain the case marking on the argument structure. In the lexical rules, the sole argument of an intransitive verb is specified to have the focus case, while arguments of transitive verbs are not specified in the lexical entry, to accommodate agent-focus and patient-focus possibilities.

The 'focus-case' system is thus the odd-one-out in two ways: On the one hand, it exercises

---

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

[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.

[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).

a logical possibility left unexploited in the other case systems. On the other hand, it was attested for only one language family. In this paper, we describe how the analysis provides a similarly elegant account for Kolyma Yukaghir (ISO 639-3: yux), a language of North-Eastern Russia.

3 Kolyma Yukaghir

Yukaghir languages 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 deviations from SOV word order are attested (Maslova, 2003, p.341). 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 five major moods (inferential, prospective, irrealis, and periphrastic prospective) (*Ibid.* p.6).

One of the most notable features of the language though 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. 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 case ('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 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.

3.2 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).[4]

---

[4]Abbreviations used in Yukaghir examples:

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

(3) tāt touke-lek jede-l
CA dog-PRED appear-SF

'Then {a dog}$_{Foc}$ appeared.' [yux] (Maslova, 2003:9)

(4) met ejre-je
I.NOM walk-1SG

'I walked.' [yux]

3.3 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 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 pronouns and the zero-marked form (equivalent to nominative) for non-pronouns (9). If the subject is 3rd person, a definite object will be accusative and an indefinite instrumental.

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

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

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

(6) tudel tolow-le kudde-m
he.NOM deer-INSTR kill-3SG

'He killed a deer.' [yux] (Maslova, 2003:10)

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

(7) met tolow kudede
I.NOM deer.NOM kill.1SG

'I killed a deer.' [yux] (Maslova, 2003:10)

(8) tudel met kønme-gele juø-m
he.NOM my friend-ACC see-3SG

'He saw my friend.' [yux] (Maslova, 2003:10)

(9) met tet-ul juø
I.NOM you-NFO see.1SG

'I saw you.' [yux] (Maslova, 2003:10)

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.

4 Analysis of Yukaghir in the Grammar Matrix

In order to test our analysis of the case system, we need our grammar fragment to handle enough other basic facts to parse our test sentences. We created the fragment through the Grammar Matrix customization system. In the process, we made a few simplifying decisions: In particular, we said that the language is simply SOV, since it works for the most basic sentences which exemplify choice of case frame in intransitive and transitive clause. We also did not implement any of the moods and only implemented habitual aspect as an example. 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 Yukaghir. 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). Thus, we called predicative the "focus case", accusative the "O-case", and nominative the "A-case".[5] 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.

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 worked on in the Morphology section, though, for a more extended analysis we would likely have more classes (for example, verbs of cognition, qualitative verbs).

4.3 Morphology

Using the focus case option requires additional work in the Morphology section of the customiza-

tion system, since, unlike other case options, focus case does not suggest that the choice of a case frame depends on a verb's class. However, picking the option does create a case frame that can be used in verb lexical rules as constraint on the verb's argument structure. In case of Yukaghir, the automatically generated "nominative-predicative" 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 (see §4.3.2).

4.3.1 Noun/Pronoun inflection rules

The key observation for the analysis for nouns and pronouns is that the NFO case marker on the pronouns is used 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). Thus, it makes sense to say the nominative zero-marker marks nouns for nominative ("A-case" in the focus case library terms) as well as for NFO. In contrast, pronouns must have separate lexical rules for nominative and NFO. Then, in the verb lexical rules section, it is sufficient to constrain the *tr-nf* object to be in NFO case. We analyze both the (surface) instrumental and accusative as marking nouns as [CASE acc].[6] This allows for the instrumental forms but does not properly 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.

4.3.2 Verb inflection rules

In the basic morphology that we consider (minus aspect, mood, and non-finite forms) the position classes are chained as follows:

(10)   NEG-stem-TENSE-AGR (Maslova, 2003:149)

However, in 3rd person plural, the tense marker is in between two markers which both mark the verb for person and number.[7] 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:

(11)   juø-ŋi-te-m
       see-3PL-FUT-3PL
       'They will see.' [yux] (Maslova, 2003:150)[8]

(12)   juø-0-te-m
       see-NON3PL-FUT-3SG
       '(S)he will see.' [yux] (Maslova, 2003:150)

(13)   ejre-ŋi-0-0
       walk-3PL-NONFUT-3PL
       'They walk.' [yux] (Maslova, 2003:150)

(14)   ejre-ŋi-te-j
       walk-3PL-FUT-3PL
       'They will walk.' [yux] (Maslova, 2003:150)

With the 3rd person plural position class in place, and with the future/non-future position class taking the output of the 3rd person plural 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 (10).

---

[7]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.

[8]Examples (11)–(14) come from a morphological table which we used for reference and were not originally in the form of IGT.

$$\begin{bmatrix} \textit{intr-sf-lex-rule} \\ \text{ARG-ST} & \left\langle \begin{bmatrix} \text{OPT} & - \\ \text{CASE} & \text{pre} \end{bmatrix} \right\rangle \\ \text{DTR} & \textit{tense-lex-rule} \end{bmatrix}$$

$$\begin{bmatrix} \textit{tr-nf-12per-lex-rule} \\ \text{ARG-ST} & \left\langle \begin{bmatrix} \text{CASE} & \text{nom} \\ \text{PER} & \text{non-3rd} \end{bmatrix}, \begin{bmatrix} \text{CASE} & \text{nfo} \end{bmatrix} \right\rangle \\ \text{DTR} & \textit{tense-lex-rule} \end{bmatrix}$$
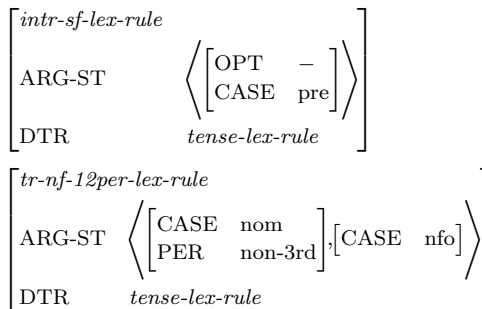
Figure 1: Sample lexical rules

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 paradigm of markers 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. 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 1 illustrates how the case frames are constrained for *intr-sf* and *tr-nf-12per*, respectively.

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.150). Furthermore, we note that this results in a working, testable grammar fragment for Yukaghir.

Coverage	Overgeneration
75.3 %	4.4%

Table 2: Yukaghir Grammar Performance

5 Evaluation

We used 175 sentences,[9] 85 grammatical, 90 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.[10]

The overgeneration is accounted for by the lack of distinction between definite and indefinite NPs in the Matrix. Both example (6) and the following ungrammatical example (15) will be parsed, though ideally we would like to reject (15):

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

As noted above, this can be handled by adding constraints on definiteness, such that the accusative marker is made incompatible with the indefinite bare NP construction.

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

6 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) 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 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. This latter can be easily added via direct editing of the grammar. 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, E. M., Drellishak, S., Fokkens, A., Poulson, L., & Saleem, S. (2010). Grammar customization. *Research on Language & Computation*, 1-50. (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). Taipei, Taiwan.

Borthen, K., & Haugereid, P. (2005). Representing referential properties of nominals. *Research on Language and Computation*, *3*(2), 221-246.

Copestake, A. (2002). *Implementing typed feature structure grammars.* Stanford, CA: CSLI Publications.

Drellishak, S. (2009). *Widespread but not universal: Improving the typological coverage of the Grammar Matrix.* Unpublished doctoral dissertation, University of Washington.

Fortescue, M. (1996). Grammaticalized focus in Yukaghir. is it really grammaticalized and is it really focus? In *Content, expression and structure:*

---

[9]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.

[10]The grammar and testsuite will be made available for download.

Studies in Danish structural grammar (pp. 17–38). John Benjamins Publishing.

Goodman, M. W. (2013). Generation of machine-readable morphological rules with human readable input. *UW Working Papers in Linguistics*, *30*.

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

Maslova, E. (2003). *A grammar of Kolyma Yukaghir*. Berlin: Mouton De Gruyter.

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

Müller, S. (2001). The passive as a lexical rule. In *Proceedings of the 7th international conference on hpsg* (pp. 247–266).

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

Pollard, C., & Sag, I. A. (1994). *Head-Driven Phrase Structure Grammar*. Chicago, IL and Stanford, CA: The University of Chicago Press and CSLI Publications.

Przepiórkowski, A. (1996). Case assignment in Polish: Towards an HPSG analysis. In C. Grover & E. Vallduvĩŋ (Eds.), *Studies in HPSG* (Vol. 12, pp. 191–228). Centre for Cognitive Science, University of Edinburgh.

Song, S. (2014). *A grammar library for information structure*. Unpublished doctoral dissertation, University of Washington.