1 *The issue of structural case*

1.1 Introduction: Some challenges of structural case assignment

Case in the linguistic sense is known to be a morphosyntactic device that helps to indicate – imperfectly, but often usefully – what role a noun phrase (NP, DP, etc.) has within a larger grammatical structure. But what kind of device is it exactly? This is a standard topic in morphosyntax, and has been studied extensively from many perspectives. For example, see Blake (2001) for empirical background, Butt (2006) for orientation to a range of theoretical approaches, and many contributions in Malchukov and Spencer (2009) for a sample of current perspectives. And yet there is still much to be done and much to understand, particularly from a perspective that tries to combine formal-generative explicitness and precision with a relatively broad typological awareness of the range of natural language phenomena ("Formal Generative Typology" in the sense of Baker [2010a]).

For example, nominative and accusative are two structural cases in Sakha, a Turkic language spoken in Siberia (also called Yakut) (Vinokurova [2005], Baker and Vinokurova [2010]). I begin with this language because it is not particularly familiar, but neither is it particularly strange, it being a reasonably typical nominative-accusative language. (I also happen to know something about it.) In a simple clause, the subject or agent is nominative, which is morphologically unmarked (there is no overt affix on the noun stem), whereas the object or theme, if there is one, bears an allomorph of the accusative suffix *-(n)I*.

(1) a. Min kel-li-m.
I.NOM come-PAST-1sS
'I came.'

b. Min oloppoh-u aldjat-ty-m.
I.NOM chair-ACC break-PAST-1sS
'I broke the chair.'

(Vinokurova [2005: 285])

c. Erel kinige-ni atyylas-ta.
Erel.NOM book-ACC buy-PAST.3sS
'Erel bought the book.'

For this type of data, it does not matter too much whether one states the case marking principles in terms of thematic roles, grammatical functions, structural positions, or some combination of the three. All versions can get the same results because the simple examples are, well, simple.

Indeed, for some cases, there might be little more to say than this from a syntactic perspective. For example, the ablative case in Sakha is not used for core arguments or grammatical functions, and it does have a fairly straightforward meaning. To a good first approximation, it is used on all and only those NPs with the meaning 'from' (Krueger [1962: 84], Stachowski and Menz [1998: 429]), as in (2).[1]

(2) Bihigi beqehee Saaska-ttan suruk tut.
we yesterday Saaska-ABL letter receive
'We received a letter from Sakha yesterday.'
(Vinokurova [2005: 241])

So we might say that the ablative suffix *–ttan* in Sakha corresponds fairly directly to the preposition *from* in English. It has a similar meaning, and it, plus the associated noun phrase, has a similar syntactic distribution – for example, as an "extra" adjunct phrase included in the larger verb phrase. We may then say that *Saaskattan* in (2) is, essentially, an adpositional phrase (PP).[2] That is perhaps nearly all there is to say about this sort of so-called semantic or

[1] As a small addendum, ablative in Sakha can also be used on causal adjuncts, like *ardaq-tan* 'because of the rain'.

[2] See, for example, Blake (2001) on the functional equivalence of semantic cases and adpositions, and the diachronic relationships between them. From a generative perspective, there are two plausible ways to work this out, which are technically different. One is to say that the morpheme *-ttan* is a direct realization of the P meaning 'from'; it appears on the noun as a result of cliticization or morphological merger (see [ia]). The other is to say that the P meaning 'from' corresponds to a null morpheme, but it assigns its own distinctive brand of oblique case to its NP complement, and that is realized as ablative, as sketched in (ib) (Emonds [1985: 224–237], McFadden [2004], also Baker and Kramer [2014] on Amharic, and others). Indeed, in some languages both the P and the case it assigns seem to be spelled out as separate morphemes on the noun, resulting in what can be described as bimorphemic case markers, like the Lezgian example in (ic).

(i) a. [Saaska+OBL FROM] → *Saaska-Ø-ttan*
b. [Saaska+ABL FROM] → *Saaska-ttan Ø*
c. [BEAR-OBL UNDER] → *sew-re-k* 'under the bear'
(Lezgian, Haspelmath [1993: 74])

Which of these analyses is to be used for which inherent/semantic cases is presumably to be decided by careful consideration of the morphological details. (For example, does the case spread onto modifers of the noun? Is the same case assigned by any other element? How does the oblique nominal compare with clearer instances of PP in the language, with overt separate P?) These questions are interesting on a local level, but typically do not have too much broad syntactic significance. Therefore, I do not consider them here.

inherent case with regard to syntax. And languages may have many inherent cases of this sort: for example, Finnish has eleven (Olli [1958: 35–36]) and Lezgian has fourteen (Haspelmath [1993: 74]).

But this is certainly not all there is to say about accusative or nominative in Sakha, which do not correspond to adpositions in English, which do not have consistent semantic values, and which seem to be used more dynamically. It is these so-called structural cases that this book is primarily about.

1.1.1 The problem of language-particular detail

The structural-grammatical cases are notably not like the inherent-semantic cases, in that they can change depending on the syntactic context. For example, the passive sentence in (3) contains a theme argument semantically comparable to the one in (1b). But in (3) this nominal is marked with the (null) nominative case, not with the accusative.

(3) Caakky aldjat-ylyn-na.
cup break-PASS-PAST.3sS
'The cup was broken.'
(B&V: 608)

Therefore, the affix *-(n)I* cannot simply be regarded as a marker of the theme-patient thematic role, the way that *-ttan* can be regarded as a marker of the source thematic role.

Conversely, the embedded sentence in (4) has an agentive subject, comparable to the one in (1a) (Vinokurova [2005: 366]). Nevertheless, in this sentence the comer is marked accusative, not nominative.

(4) Keskil [Aisen-y kel-bet dien] xomoj-do.
Keskil Aisen-ACC come-NEG.AOR.3sS that become.sad-PAST.3sS
'Keskil became sad that (because) Aisen is not coming.'

Examples like these show that one cannot state (nearly) exceptionless rules that relate these morphological markings to thematic roles like agent and patient-theme in Sakha. Indeed, one cannot state them for structural cases in most other languages either. In functionalist terms (Malchukov and De Swart [2009], Seiwierska and Bakker [2009]), the *indexing* (or *characterizing*) role of these structural cases is much less clear.

One might then switch to stating the rules of structural case marking in terms of grammatical functions like subject and object, rather than in terms of thematic roles. In these terms, a noun phrase is nominative if it is the subject of the clause, and accusative if it is the object. This type of formulation might work for (3) as well as for (1), assuming that the object of a transitive sentence corresponds to

the subject of the passive version. But it is less clear that it works for (4). This would only work if one said that (4) (like so-called exceptional case marking in English), is an instance of "raising to object." But there is little motivation for this in Sakha, apart from the case marking. Note that the matrix verb in (4) is an intransitive one, 'become sad', which is not the sort of verb that one would expect to take an object, thematic or otherwise. Indeed, it bears the anticausative suffix *-j*, which otherwise marks intransitive verbs of the unaccusative class; see Baker and Vinokurova (2010: 617–618) for discussion and further evidence.

Another problem for equating structural case directly with grammatical function is that the object in a sentence like (5) is not accusative, but rather nominative/unmarked, in contrast to (1c). This is true despite the fact that 'book' is clearly not the grammatical subject here, but rather *Erel* is, as confirmed by subject-verb agreement, subject-object-verb word order and other considerations. (This is an instance of so-called differential object marking (DOM); see Aissen [2003], among many others; (2) is another example.)

(5) Erel kinige atyylas-ta.
Erel book buy-PAST.3sS
'Erel bought a book/books.'
(Vinokurova [2005: 322])

So the structural case of an NP is not a direct function of that NP's independently determined grammatical function, any more than it is a direct function of its thematic role.

Perhaps then we should use structural terms instead of thematic roles and grammatical functions to formulate the principles of case distribution. Indeed, in this work I claim (non-uniquely) that this is essentially the correct approach. But it is not an easy or trivial approach, because the structural differences can be subtle. For example, there is no gross syntactic difference in the position of the theme/object between (1c) and (5); if anything, the superficial difference appears to be a semantic one, whether the object is interpreted as a nonspecific indefinite ('some book(s)') or as specific ('the book' or 'a certain book'). However, a structural difference comes to light when an adverb is included. The bare object with a nonspecific indefinite interpretation must be immediately before the verb, whereas the accusative object with a specific or definite interpretation need not be – indeed prefers not to be, as seen in (6).

(6) a. Masha türgennik salamaat-(y) sie-te.
Masha quickly porridge-ACC eat-PAST.3sS
'Masha ate porridge quickly.' (ACC on 'porridge' only if it has contrastive focus)
(B&V: 602)

b. *Masha salamaat türgennik sie-te.
Masha porridge quickly eat-PAST.3sS
'Masha ate porridge quickly.'
c. Masha salamaat-y türgennik sie-te.
Masha porridge-ACC quickly eat-PAST.3sS
'Masha ate the porridge quickly.'

Another sign that subtle, arguably structural differences can influence case marking in Sakha is the fact that the theme argument in a passive clause can be accusative rather than nominative (see (3)). Indeed, the theme argument must be accusative if agent-oriented adverbs like 'intentionally' and 'with a hammer' are present, as shown in (7).

(7) Caakky-ny sorujan ötüje-nen aldjat-ylyn-na.
cup-ACC intentionally hammer-INST break-PASS-PAST.3sS
'The cup was intentionally broken with a hammer.' (* with *caakky* 'cup.NOM')
(Vinokurova [2005: 336])

Baker and Vinokurova argue that the agent-oriented adverbs in (7) imply that there is a covert agent in the syntactic representation of the clause in (7), but not in (3), and this covert agent influences the case marking on the theme.

A third sign that structural differences influence case in Sakha is that the subject of an embedded clause may be nominative as well as accusative, as in (4). Indeed, the embedded subject must be nominative if it follows an adverb that modifies the lower verb, as shown in (8b), as opposed to (8a) (B&V: 615–616).

(8) a. Min ehigi/ehigi-ni bügün kyaj-yax-xyt dien erem-mit-im.
I you/you-ACC today win-FUT-2pS that hope-PAST-1sS
'I hoped that you would win today.'
b. Min [sarsyn ehigi-(*ni) kel-iex-xit dien] ihit-ti-m.
I(NOM) tomorrow you-(*ACC) come-FUT-2pS that hear-PAST-1sS
'I heard that tomorrow you will come.'

This range of data shows us two things. The first is that syntactic structure has the potential to explain fine-grained differences in structural case marking that cannot be explained purely in terms of thematic role or simple grammatical function. The term "structural case" is thus not a misnomer, but points toward an important truth. The second is that it will be none too easy to get an account even in structural terms. The syntax will have to be fairly detailed to distinguish (1c) from (5), (3) from (7), and (8a) from (8b). It may not be immediately obvious how to get a unified syntactic account of these three differences, which may not seem to have much to do with each other.

Part of the challenge of structural case, then, is that it is easy to get principles of case assignment that *sort of* work, but it is hard to get ones that work *exactly*, over a broad domain in a particular language. Nor is Sakha notably more difficult than other languages in these respects. There is nothing unique to my framing of this problem; it has been a classic problem in syntactic theory for years. But it is not a solved problem. In this work, I attempt to take a big step forward in solving it.

1.1.2 The problem of crosslinguistic generality

The issue of structural case gets even harder and more interesting when it is given a crosslinguistic dimension, within a theory that has universal aspirations. We have seen that structural details matter in Sakha. They also matter in other languages, and they matter differently.

I chose a less familiar language for my initial presentation in the hope that many readers would be struck by both similarities with and differences from languages they already know. For example, there are many languages with data like (1) in Sakha, in which the object of a transitive verb is distinguished from the subject of a transitive or intransitive verb by bearing a morphological marker. They include Turkish, Tamil, Amharic, Korean, Quechua, Hopi, Russian – and even English, when one considers the differing forms of some personal pronouns. But when it comes to examples like (3)–(8), one notices unfamiliar details. For example, English has a passive, but the theme argument of a monotransitive passive must be nominative, never accusative, even in the presence of agent-oriented adverbs.

(9) a. He was beaten on purpose with a hammer.
b. *Him was beaten on purpose with a hammer.

Similarly, in English the subject of an embedded clause can be marked accusative, but only if the clause is nonfinite ((10a) versus (10b)), whereas the embedded clause is finite in (8a) from Sakha. Also the embedded clause must be a complement of the matrix verb in English, whereas it can be an adjunct in Sakha, as seen in (4). This type of accusative case marking is also possible with a smaller range of matrix verbs in English than in Sakha, so (10b) is not very good with the verb *hope* (cf. (8a)).

(10) a. I hoped/expected that she (*her) would win today.
b. I expected/??hoped her to win today.

Indeed, it is notable that Turkish, although it is historically related to Sakha, is more like English than like Sakha in these details (George and Kornfilt

[1981]). So we can have significant differences in the grammar of structural case marking within a family, and significant similarities in the grammar of case marking across families.

Turkish is like Sakha in that some objects are marked accusative and others not ((1b,c) versus (5)). But not all languages with overt accusative marking on common nouns are like this. Cuzco Quechua (CQ), for example, is not: in its matrix clauses, objects are marked with overt accusative case even if they are nonspecific indefinites adjacent to the verb (Liliana Sanchez, personal communication, 2012):[3]

(11) Juan wawakuna-man miski-*(ta) qunpuni.
Juan children-DAT candy-ACC give.3S.HAB
'Juan gives candy to the children (habitually).'

In terms of the recent literature, Turkish and Sakha are differential object-marking (DOM) languages (Aissen [2003]), but Quechua is not.

So we find ourselves in a familiar kind of quandary. We want to capture the similarities across languages that characterize a system of (say) accusative case marking. But we also need to capture the differences. This raises such questions as what is the core of the notion of accusative case marking (if any), and what is its range of allowable variation? This is the classic Principles and Parameters question (Chomsky and Lasnik [1993]) applied to this particular empirical domain – a domain that is relatively finite and accessible, with much crosslinguistic information available, and one that is interesting and strategic for our understanding of grammar as a whole.

There are also larger-scale differences among languages when it comes to case marking. It is well known that not all languages with overt case marking have a nominative-accusative system, where there is a special case marker for the direct object of a transitive clause. Famously, there are also ergative languages, in which a special affix marks the subject of a transitive clause, while the subject of an intransitive clause and the object of a transitive clause have the same marking (often null) (Comrie [1978], Dixon [1979], Blake [1994], Dixon [1994]). (12) shows a pattern of this kind in Shipibo, a language from the Panoan family, spoken in Peru (Valenzuela [2003], Baker [2014a]).

[3] See also Cole (1985: 70–71) on Imbabura Quechua. Some embedded clauses in CQ are different, in that accusative case can be or must be omitted on the object, because the clauses are nominalized (Lefebvre and Muysken [1988]). I discuss this briefly in sections 4.1.3 and 5.3.

(12) a. *Maria-nin-ra* ochiti noko-ke. (Shipibo)
Maria-ERG-PRT dog find-PRF
'Maria found the dog.'
b. Maria-ra ka-ke.
Maria-PRT go-PRF
'Maria went.'

Indeed, ergative languages are not much less common than accusative languages among languages with overt structural case markers (e.g. 32 ergative to 46 accusative languages out of 190 total languages in Comrie [2005]).[4] Some rarer alignment types are also known to exist: tripartite languages, in which intransitive subjects, transitive subjects, and transitive objects are all marked differently (4 out of 190), and marked nominative languages, in which it is the subject of the clause (transitive or intransitive) that bears an overt affix rather than the object (6 out of 190). This work attempts to account for these seemingly larger-scale differences in case marking as well as the smaller-scale differences.

A final piece of the introductory puzzle is that even languages that seem to have quite different kinds of case system can show surprising similarities when one considers details of the system. For example, Shipibo has a small number of verbs that take two distinct NP arguments, neither of which is marked ergative; rather, both NPs are absolutive in (13) (see Valenzuela [2003: 339, 342–344], Baker [2014a]).

(13) Jose-ra yapa keen-ai.
José-PRT fish want-IMPF
'José wants some fish.'

These verbs have psychological meanings, where the subject is an experiencer or possessor, rather than an agent. Korean is quite different from Shipibo in that it is a nominative-accusative language. But it also has a minority pattern in which both arguments have the same case, namely nominative. Moreover, it is nonagentive predicates with experiencer subjects that have this special behavior in Korean, just as in Shipibo.

[4] It is well known that languages can be ergative in different senses. In some, the subject of a transitive verb is distinguished from the subject of an intransitive verb and the object of a transitive verb only by case marking and/or agreement (morphological ergativity). In others, there are also syntactic processes that treat transitive subjects in a distinctive way (syntactic ergativity: see Anderson [1976], Marantz [1984], Dixon [1994]). Putting aside the special case of Dyirbal, current literature shows that syntactic ergativity concerns primarily A-bar extraction phenomena (Deal [in press], Polinsky [in press]). Since this book is about case marking, I will be concerned with case marking in ergative languages of both types, but what distinguishes them is not investigated. In practice, Shipibo and the other ergative languages I have studied most are morphologically ergative.

(14) a. Chelswu-ka sangca-lul mantul-ess-ta. (normal transitive)
Chelswu-NOM box-ACC make-PAST-DEC
'Chelswu made a box.'
(Koak [2012: 58])

b. John-i Mary-ka mwusep-ta. (experiencer predicate)
John-NOM Mary-NOM be.afraid-DEC
'John fears Mary.'
(Levin [2013: 2])

These two languages thus have something in common, in that their special case for transitive clauses (accusative or ergative) is not used with certain experiencer predicates – a similarity that cuts across the distinction between ergative and accusative languages.

Another cross-cutting comparison of this type can be made between the ergative language Eastern Ostyak and accusative Sakha. We saw above that in Sakha when the object is a nonspecific indefinite next to the verb it is not marked for accusative case (see (5)). Something analogous happens in Eastern Ostyak: when the object is a nonspecific indefinite next to the verb, the subject is not marked for ergative case. This is shown in (15a), as compared with the normal ergative clause (15b) (Gulya [1966: 135]).

(15) a. *Mä* t'əkäjəɣlämnä ula mənɣäləm.
We.dual(NOM) younger.sister.COM berry pick.PAST.1pS
'I went to pick berries with my younger sister.'

b. *Mə-ŋən* ləɣə əllə juɣ kanŋa aməɣaloɣ.
We-ERG them large tree beside put.PAST.3pO/1pS
'We put them (pots of berries) beside a big tree.'

Here too we can discern something significant that Sakha and Eastern Ostyak have in common, that transcends the fact that one language is accusative and the other is ergative. Similarities like these suggest that ergative and accusative are not radically different systems, with very different principles and logics, but rather variations of a single abstract system. That is why the same kinds of factors – factors like the agentivity of the subject and the specificity of the object – can be relevant to both. We see, then, a complex pattern of both difference in the midst of similarity and similarity in the midst of difference. This is what I aspire to say something about.

1.1.3 The goals of the inquiry

A key feature of this book, then, is that it attempts to address both the problem of language-particular detail and the problem of crosslinguistic variation in a balanced way. Of course, it is impossible to do this fully in one go: one cannot

go both deeper and broader to the fullest degree in one book of limited length, even if one had all the expertise needed to do so. But there is some value in trying to advance simultaneously along both dimensions, given that some of the interesting crosslinguistic differences – and also some of the interesting crosslinguistic similarities! – only appear when one reaches a certain level of detail. Therefore, this book follows what I have called "The Middle Way" (Baker [2010a]). It is built on a study of a medium number of languages in a medium amount of detail. Roughly, this has amounted to considering some twenty languages from different families to the extent of studying at least one complete grammar of the language (not just the obvious pages on case marking) or a series of articles (not just one article), trying to take into account most of what comes up regarding structural case. The languages I have chosen to focus on are listed in (16), sorted according to their alignment types. Languages I have done direct fieldwork on are in italics.

(16)
a. Accusative languages: *Sakha*, *Tamil*, *Amharic*, Cuzco Quechua, Korean, Finnish
b. Ergative languages: *Shipibo*, Burushaski, Chukchi, Lezgian/Ingush, Greenlandic; also, more briefly, Kewa and Wardaman
c. Tripartite languages: Nez Perce, Coast Tsimshian, Semelai, Diyari/Warlpiri
d. Marked nominative languages: Choctaw, Oromo, Tukang Besi, Maricopa/Mojave (also, arguably, Korean)
e. Marked absolutive language: Nias

I mention some of these languages less often than others in this book – and sometimes I cite data of special interest from other languages, like Eastern Ostyak – but the core ideas have been developed in the laboratory of trying to construct relatively complete analyses of structural case marking for this selection of languages.

A related goal is that, for each language considered, I seek principles of case assignment that are as unified as possible. This means that, for a language like Sakha, I seek one rule of accusative case assignment that captures when NPs are accusative and when they are not over the entire range that accusative is used in in the language – and similarly for ergative, nominative, absolutive, dative, and genitive. At least that is the ideal I aim for.

Some readers will question whether this is the right goal. (Sometimes I question it myself.) Descriptive grammars typically do list a variety of disparate-seeming uses of (say) accusative case. Modern theories along the