Case study: the rise of ergativity

English, like a majority of the world’s languages, is an *accusative* language. That is, the subjects of intransitive verbs and the subjects of transitive verbs are treated identically for grammatical purposes, while direct objects are treated differently. A simple example of this is the case marking on pronouns:

i. She smiled.
ii. She saw me.
iii. I saw her.

Here the female pronoun takes the case-form *she* when it is an intransitive subject i or a transitive subject ii, but the form *her* when it is a direct object iii.

A sizeable number of languages exhibit a quite different pattern: intransitive subjects and direct objects are treated identically, while transitive subjects are treated differently. Basque is a good example: NPs like *gizona* ‘the man’ and *neska* ‘the girl’ can take either the case-ending zero or the case-ending *-k*, as follows:

iv. *Gizona heldu zen.*
   ‘The man arrived.’
v. *Gizonak neska ikusi zuen.*
   ‘The man saw the girl.’
vi. *Neskak gizona ikusi zuen.*
   ‘The girl saw the man.’

Languages such as Basque are called *ergative* languages. You can see that only a transitive subject takes the case-ending *-k*; both intransitive subjects and direct objects take the ending zero. In Basque, ergative morphology is thoroughgoing: ergative case-marking applies in all circumstances, and verbal agreement is ergative as well.

When European linguists first encountered ergative languages like Basque, they were quite bewildered by them, and they put forward quite a number of confused misinterpretations, some of which may still be encountered in popular books written by non-linguists. Today we have information on hundreds of ergative languages, and we no longer regard them as particularly exotic. The study of such languages has turned up a number of interesting points. One of these is that many ergative languages exhibit a *split* in their grammar: that is, they are ergative only in certain circumstances. One of the commonest splits is this: ergative morphology appears only in the perfective aspect or in the past tense, accusative morphology being found elsewhere. Here is an example from Pashto, a East Iranian language; the label Nom(inative) denotes the ordinary subject case, while Obl(ique) denotes the ordinary non-subject case. First, in the present tense:

vii. *ze de winem*
    I-Nom he-Obl see-I
    ‘I see him’
viii. *day maa wini*
   he-Nom I-Obl see-he
   ‘He sees me’

ix. *te maa winee*
   you-Nom I-Obl see-you
   ‘You see me’

Now, in the past tense:

x. *maa day wulid*
   I-Obl he-Nom saw-he
   ‘I saw him’

xi. *taa ze wulidem*
   you-Obl I-Nom saw-I
   ‘You saw me’

Observe the differences. In the present tense, the subject is in the subject case and the verb agrees with it, while the object stands in the oblique case. But, in the past tense, it is the object that stands in the subject case, and the verb agrees with it, while the subject is in the oblique case.

How does ergativity arise? There is one obvious possibility. Early linguists often mistakenly described ergative constructions as ‘passives’. As it happens, they are demonstrably not passives, but of course they might be *derived* from earlier passives. Suppose, in the Pashto case, that a sentence like x once meant, not ‘I saw him’, but rather ‘He was seen by me’. Its structure would then have been, literally, ‘By-me he was-seen’. This would explain at once why *day* ‘he’ is in the subject case (it was the subject) and why the verb agrees with it (verbs agree with subjects in Pashto), and also why *maa* ‘I’ is in a non-subject case (it wasn’t the subject, but an oblique noun phrase of agent). So possibly the Pashto ergative construction derives from what was formerly a passive.

This looks plausible, but, of course, it doesn’t explain why the ergative should have arisen only in the past tense. We would therefore like to have some historical information. Now, for the majority of ergative languages, we unfortunately lack any significant historical documentation, but there are exceptions. One of the major exceptions is the Indo-Iranian family of languages, to which Pashto belongs. For this group of languages we are lucky enough to have texts going back several thousand years, and so we can see the ergative construction developing over many centuries.

Consider Sanskrit, the ancient language that is the ancestor of most north Indian languages like Hindi and Bengali, most of which are also ergative today. In Sanskrit, we really do find what appears to be a passive construction, and it was indeed mostly used in the past tense. Here Nom(inative) is the ordinary subject case, Acc(usative) is the ordinary object case, and Obl(ique) is an oblique case; *Ram* is a male name, while *Sita* is a female name. In the present tense:

xii. *rāmah sītām prcchati*
    Ram-Nom Sita-Acc asks-3Sg
    ‘Ram asks Sita’
And in the past tense:

xiii. \(\text{rāmena sītā prṣṭā}\\)
\(\text{Ram-Obl Sita-Nom was-asked-Female}\\)
\(\text{‘By Ram Sita was asked’, i.e., ‘Ram asked Sita’}\\)

Word order was rather free in Sanskrit, and the subject did not have to come first. Most linguists have concluded, therefore, that the ergative constructions of the modern Indo-Iranian languages derive from original passives. At least in the past tense, these passives apparently came to be used more and more frequently, until they finally became the normal (unmarked) form, while the original active construction became so rare that it dropped out of the languages altogether. Since the passive no longer contrasted with an active, it was then reinterpreted as being itself an active construction, as it is today, but it still kept its original morphology, producing the ergative constructions we see today.

If it is right, this account illustrates two of the general pathways of syntactic change discussed above. First, there was a shift of markedness, as the formerly marked passive became so frequent that it turned into the unmarked form, while the formerly unmarked active became first highly marked and then obsolete. Second, there was a reanalysis, as the former passive construction was reanalysed as an active. It might even be possible to see events of this type as restructurings, although they seem too gradual for this.

But why should the passive have become so frequent in the first place? We don’t know, but one possible answer is politeness. Passives, being indirect and often impersonal, may often seem more polite and less abrupt than actives: compare the English passive \(\text{John’s arm has been broken}\\) with its corresponding active \(\text{Fred has broken John’s arm}\\). Such developments are attested: for example, in Malagasy, the chief language of Madagascar, in which blunt statements are generally regarded as socially inappropriate, actives, although possible, are not at all frequent. Instead passives and circumstantialis (a voice almost impossible to translate into English, emphasizing neither the agent nor the patient) are the norm.

But passives are not the only possible source of ergatives. Since ergative constructions are so often confined to the perfective aspect, we might wonder whether ergatives can be directly derived from perfective constructions, and there is good evidence that this does happen.

Recall the English perfect discussed above, in which what was originally a stative construction of the form \(\text{I have a window broken} (= ‘One of my windows is broken’)=\) was reinterpreted as a non-stative perfect with the meaning \(\text{I have broken a window} (= ‘I broke a window, and it’s not fixed yet’)=\). Note carefully that the stative construction was originally a possessive construction, essentially identical to \(\text{I have a dog}\\). It is known that possesives are a common source of statives, and hence of perfects. In English, this development has not led to ergativity, because English uses the transitive verb \(\text{have}\\) to express possession. But lots of languages have no such verb.

So how do they express possession? They do it like this (Welsh \(y\) is a grammatical particle):

xiv. \(\text{Welsh:}\\)
\(\text{Y mae gardd gennyf i}\\)
Prt is garden with me
‘A garden is with me’, i.e., ‘I have a garden’

xv. Russian:
\[ \text{U menja kniga} \]
at me book
‘A book [is] at me’, i.e., ‘I have a book’

xvi. Early Latin:
\[ \text{Est mihi liber} \]
is to-me book
‘A book is to me’, i.e., ‘I have a book’

xvii. Fijian:
\[ \text{saa ti’o vei au e dua a pua’a} \]
Asp be-at to me 3Sg one Art pig
‘A pig is to me’, i.e., ‘I have a pig’

Constructions like these are very widespread. Observe that the thing possessed stands in the subject case and takes any verbal agreement going, while the possessor stands in an oblique case. Now, suppose such a language were to use its possessive construction to create a stative, and consequently a perfect, just as in English. The result would be perfect constructions with meanings like ‘I have broken a window’, but with subject-case marking on the object, verb agreement with the object, and oblique marking on the subject. Just such an origin has therefore been proposed for ergatives in a number of languages, especially in those in which ergativity is confined to the perfective aspect. Indeed, some linguists prefer to derive the ergative constructions of the Indo-Iranian languages from an earlier perfect that was itself obtained from a possessive construction; this has the advantage of explaining why, in these languages, ergatives are usually confined to perfect aspect or to past tense (past tenses are themselves often derived from earlier perfects).