W. D. C. de Melo
Week 2, Hilary term 2007

Linguistic Typology: Core Constituents

Verbal valency

Every language has nouns and verbs. Verbs can express a variety of actions and states, but normally they have to be combined with nouns in order to do so. Valency is about how many nouns or pronouns a verb needs to express an action or state. If a verb needs two nouns or pronouns, we say that it takes two arguments / has a valency of two / is a two-place verb.

- One-place verbs: *cough*, *grin*, *sleep*
- Two-place verbs: *despise*, *hit*, *love*
- Three-place verbs: *give*, *hand over*

In addition, there are 0-place predicates (*rain*, *snow*, *thunder* and similar processes). In some languages they take dummy-subjects, in others they do not:

(1) *It is raining.*

(2) *Plu-it.* (Latin)
rain-3SG.PRES.IND.ACT
'It is raining.'

In English you need a dummy-pronoun, in Latin it would be ungrammatical; note that the third-person singular ending is simply the default ending for finite verbs.

Note also that many verbs can have alternative frames; for example, *drive* and *sing* can take objects (*a car*, *a song*), but they need not do so.

Semantic (macro-)roles

Driving and singing are completely different activities. When you drive, you use your hands and feet; when you sing, you use your mouth. Nevertheless, people who are driving and people who are singing are usually marked in the same way in a language, for instance by placing them in front of the verb or by putting them in the nominative case. It would be inefficient if you had to learn different word orders or different cases for every semantic role because there are as many semantic roles as there are verbs.

Languages always lump several semantic roles together to form macro-roles. Macro-roles are more or less the same for every language. We can distinguish between:

a) agent:

(3) ***John*** *wrote a book.*

b) patient:

(4) *John packed* ***the Christmas presents****.*

c) recipient:

(5) *John gave the presents* ***to his wife****.*

d) beneficiary:

(6) *John wrote a cheque* ***for his employee****.*

e) instrument:

(7) *John's wife opened the presents* ***with a knife****.*

f) location:

(8) *You will find the bread* ***in the basket****.*

g) time:

(9) *You will have a break* ***at lunchtime****.*

The subject

In what follows, I adopt a semantic definition of subject similar to that used in a Functional Grammar framework:

The subject is that constituent which is the vantage point from which an event is looked at.

(10) ***John*** *ate all the sandwiches.*

(11) ***All the sandwiches*** *were eaten by John.*

Exx. 10 and 11 present the same event, but from different perspectives. In Ex. 10 the event is presented from John's perspective, so he is the subject. In Ex. 11 the event is presented from the perspective of the sandwiches; they are the subject. It is more normal to present this event from John's perspective. If you choose a different perspective, you either need a different verb or a different voice. In English this alternative voice is the passive, which is much rarer than the active (between 1% in spontaneous narratives and 18% in academic prose).

Subject assignment (in the unmarked active voice) is not random:

- If a verb has only one constituent, this will be the subject.

- If a verb has more than one constituent, subject assignment follows this (incomplete) hierarchy:

  animate agent > inanimate agent / instrument > marginally affected patient > highly affected patient

  The more to the left a constituent is, the more likely it is to become subject.

Thus, if a door (patient) is opened and John is the agent and a key is the instrument, the unmarked construction will be:

(12) *John* (agent = subject) *opened the door* (patient) *with a key* (instrument).

Object assignment

If the subject is the primary vantage point from which an event is viewed, the (direct) object could be described as the secondary vantage point. If you take a two-place verb, the constituent which is not subject will be object. If you take a three-place verb and assign a subject, the direct object will be the second-most important constituent, that constituent which is more highly affected by the action than the other.

(13) *John sprayed the wall* (direct object) *with the paint.*

(14) *John sprayed the paint* (direct object) *on the wall.*

Ex. 13 implies that the wall is completely covered with paint because as the direct object it is highly affected; Ex. 14 carries no such implication, but implies that the paint has been used up, which is not necessarily implied in Ex. 13.

A, P, and S

In typological studies, the subject of a two-place verb is abbreviated A and the object, P (or occasionally O). The subject of a one-place verb is abbreviated S.

Note that A and P are assigned on a semantic basis; the more agentive macro-role becomes A and the more patient-like macro-role becomes P. S is assigned without any semantic considerations:

(15) *John ran.*

(16) *John collapsed.*

The S in Ex. 15 is an agent, the S in Ex. 16 is a patient.

A, P, and S can be marked by word order (English), by cases (Ancient Greek), by adpositions (Japanese), or by a combination of these.

Just as many semantic roles combine to give a handful of macro-roles, various languages mark A and S or P and S in the same way. We shall now look at such systems.

Nominative-accusative systems

Nominative-accusative systems mark A (nominative) differently from P (accusative), and S is marked in the same way as A, regardless of whether it is an agent or a patient.

Most European and most African languages follow this pattern (exceptions are rare, e.g. Basque).

Example language: German

In German, articles and to some extent nouns are marked for case. In main clauses the most frequent order is Subject – Verb – Object (if any).

(17) *Der* *Mann kauft* *den* *Tisch.*
the.MASC.NOM. man buy-3SG the.MASC.ACC table
'The man is buying the table.'

(18) *Der* *Mann läuft.*
the.MASC.NOM. man run.3SG
'The man is running.'

(19) *Der* *Mann schläft.*
the.MASC.NOM. man sleep.3SG
'The man is sleeping.'

In Ex. 17 A, *der Mann*, is in the nominative. P, *den Tisch*, is in the accusative. In Exx. 18 and 19 *der Mann* is also in the nominative; in Ex. 18 the noun phrase is an S in agent function and in Ex. 19 it is an S in patient / undergoer function.

Ergative-absolutive systems

Ergative-absolutive systems mark A (ergative) differently from P (absolutive), and S is marked in the same way as P, regardless of whether it is an agent or a patient.

Most Australian languages and many South American languages follow this pattern; in Europe it is found in the language isolate Basque, and among Indo-European languages it is wide-spread in the Indic branch.

Example language: Dyirbal (R. M. W. Dixon (1972), *The Dyirbal Language of North Queensland*, Cambridge)

In this section we shall restrict ourselves to nouns. For pronouns see below. Both nouns and their classifiers are marked for case. The word order is relatively free.

(20) *Balan* *ɖugumbil* *baŋgul*
CLASSII.ABS woman.ABS CLASSI.ERG
yaɽa-ŋgu *balgan.*
man-ERG hit.NONFUT
'The man is hitting the woman.'

(21) *Bayi* *yaɽa* *ɲinaɲu* *diban-da.*
CLASSI.ABS man.ABS sit.NONFUT stone-LOC
'The man sat on the stone.'

Sentence 20 is transitive; it has an ergative A and an absolutive P. Sentence 21 is intransitive, so S is in the unmarked absolutive case.

Morphological ergativity and syntactic ergativity

English is a typical nominative-accusative language. If two English clauses are conjoined, the second subject need not be expressed if it is identical with the first, regardless of whether these subjects are both S's, both A's, or A in one clause and S in the other. The following examples illustrate this:

(22) *John slept and (he) snored.*
S (S)

(23) *John kicked the ball, but (he) missed the goal.*
A P (A) P

(24) *John walked and (he) sang a song.*
S (A) P

(25) *John sang a song and (he) walked.*
A P (S)

In Ex. 22, the second S is coreferential with the first S, so the second one can be left unexpressed; in Ex. 23, the second A can be left out because it is coreferential with a preceding A; in Ex. 24, equi-NP deletion of A is possible because there is coreference with S; and in Ex. 25, S can be deleted because there is a preceding A with the same reference.

Most languages with ergative-absolutive morphology function on a nominative-accusative basis syntactically, i.e. they have the same co-ordination patterns as English. Some languages with ergative-absolutive morphology are also syntactically ergative. In Dyirbal you get co-ordination patterns like this:

(26) *John met Mary, but Jim asked (her) out.*
A P A (P)

(27) *John was sleeping at his desk and his boss saw (him).*
S A (P)

(28) *Mary was observing John and (he) felt ill at ease.*
A P (S)

Equi-NP deletion is possible if the coreferential NP's are P and P, S and P, P and S, or S and S.

Note that there do not seem to be any languages which are morphologically nominative-accusative, but syntactically ergative-absolutive.

Morphological split ergativity

Some languages have both nominative-accusative and ergative-absolutive marking.

If this split has to do with tense and aspect, it is always the past / perfective which is marked for ergative-absolutive and the non-past / imperfective which is marked for nominative-accusative. The reason for this seems to be that past tenses typically focus on the patient rather than the agent. In languages that are purely nominative-accusative, the passive is more frequent in the past tenses than in the present or future, which also shows that the patient is more important in the past. It seems that the ergative-marking in the past tenses in the Indic branch of Indo-European in fact arose by reanalysis of passives.

If the split has to do with NP types, it follows Silverstein's hierarchy:

1st / 2nd person pronouns > 3rd person pronouns > proper names and kinship terms > human common nouns > other animate common nouns > inanimate common nouns,

where x > y means that x is more likely to exhibit nominative-accusative morphology. The hierarchy is also referred to as the 'animacy hierarchy' and the NP's to the left are regarded as more 'animate', but note that the hierarchy is not just about animacy, but also about topicality and frequency.

This type of split can be observed in Dyirbal:

	S	A	P
we (dual)	*ŋali*	*ŋali*	*ŋali-ɲa*
rainbow	*yamani*	*yamani-gu*	*yamani*

Marking of core cases among nouns and pronouns

The pronouns have nominative-accusative marking, the common nouns have ergative-absolutive marking.

The explanation for this type of split is the economy principle. The ergative is often formed by adding a morpheme to the unmarked absolutive, just as the accusative is often formed by adding a morpheme to the unmarked nominative. If you do frequency counts, you will see that across languages the NP's at the left end of the hierarchy occur most frequently in A or S function, but rarely in P function. In order to mark as little as possible, you need nominative-accusative marking. NP's at the right end of the periphery occur most frequently in P or S function, but rarely in A function. In order to mark as little as possible, you need ergative-absolutive marking.

Tripartite systems

Tripartite systems, with three different markers for A, P, and S, are uneconomical and hence rare as a general marking strategy. They occur relatively frequently as a partial marking strategy if the language has animacy-based split marking. NP's at the left end have nominative-accusative marking, with no case morpheme for A and S, but an accusative marker for P. NP's at the right end have ergative-absolutive marking, with no case morphems for P and S, but an ergative marker for A. NP's in the middle have no case marker for S, but an ergative marker for A and an accusative marker for P.

Excursus: comments on transitivity

Earlier on I introduced the distinction between transitive and intransitive. Actually, the distinction is not always easy to draw. There are of course prototypically transitive verbs like *hit* or *squeeze*, which have fully affected patients. And there are prototypically intransitive verbs like *sit* or *sleep*. But there are also many verbs in between these two poles, for instance *like* or *wait for*. Here the object is not fully affected. There are different strategies a language can employ. English marks the object of *like* in the same way as the object of *hit*, i.e. as a fully affected patient. But note that the object of *wait* is a prepositional one, which indicates reduced transitivity. There are a number of factors apart from verbal semantics which can reduce transitivity: negation, irrealis mood, imperfective aspect etc. In nominative-accusative languages, reduced transitivity can only be indicated among the objects. In Hungarian, objects are in the accusative if they are completely affected, but in the genitive if they are not completely affected. In ergative-absolutive languages, reduced transitivity can be marked among subjects as well as objects. Compare the following data from Djaru (Pama Nyungan, western Australia; data from T. Tsunoda (1981), *The Djaru Language of Kimberley, Western Australia*, Canberra):

- fully transitive: 'hit' or 'eat', ergative subject, absolutive object:

  (29) *Mawun-du guɲar buŋ-an.*
  man-ERG dog hit-PRES
  'A man hits a dog.' (p.97)

- semi-transitive: 'look for' or 'wait for', ergative subject, dative object:

  (30) *Mawun-du ŋa-la ɖaɖi-wu muwuwuŋ-an.*
  man-ERG C-3SG.DAT kangaroo-DAT1 search-PRES
  'A man looks for a kangaroo.' (p.97)

- semi-intransitive: 'talk' or 'play', absolutive subject, absolutive object:

  (31) *Mawun ɖaɾu maɳ-an.*
  man Djaru talk-PRES
  'A man talks Djaru.' (p.97)

- fully intransitive: 'go' or 'sit', absolutive subject, no object:

  (32) *Mawun jan-an.*
  man go-PRES
  'A man goes.' (p.97)

This shows clearly that transitivity and intransitivity are merely the endpoints on a scale.

Neutral systems

Neutral systems, with no distinction between A, P, and S anywhere, not even in terms of word order, are extremely rare. Neutral systems do, however, occur as subsystems: if a language marks highly 'animate' NP's as nominative-accusative and highly 'inanimate' NP's as ergative-absolutive, there may be some NP's of intermediate 'animacy' which have no marking at all.

Active-inactive systems

Active-inactive systems pay more attention to semantic factors than others. A and P are marked differently, and S is marked like A if it is an agent, but like P if it is a patient.

Example language: Mohawk (Iroquoian, Canada)

Two-place verbs:

(33) *i**k**shere?s* 'I chase it'

(34) ***wák**shere?s* 'it chases me'

→ *-k-* = first person, agent; *wak-* = first person, patient

One-place verbs, S = agent:

(35) ***k**atáwas* 'I swim'

One-place verb, S = patient:

(36) ***wak**óre?sa?* 'I'm fat'

It is not clear if the categories subject and object are universal; for an active-inactive language they seem unnecessary, as it is macro-roles which are marked rather than perspectives. Active-inactive languages generally have no alternative voices like passives.

The type of marking among one-place verbs may be completely determined by the verb or by factors such as definiteness, animacy etc. But sometimes there is also a certain amount of flexibility, e.g. 'I cough intentionally' (S marked as agent) vs. 'I cough (and can't help it' (S marked as patient).

Inverse systems

I can merely mention the existence of 'inverse systems' here. They are relatively wide-spread among North American languages, irrespective of language family. The animacy hierarchy is important here. Two-place verbs take two arguments. If the more 'animate' argument is the agent and the less 'animate' argument is the patient, the verb has no special marker; if it is the other way round, the verb takes 'inverse' marking. Again it is semantic roles that matter rather than perspectives, so subject and object may be categories that do not matter here. And again there are no alternative voices like passives.