1. Cross-linguistic definition of the notion “subject”

- Subjects in typology:

Some “prototypical” properties of subjects (based on Keenan 1976):

- **Semantics**
  The subject is typically associated with the semantic roles of AGENT, EXPERIENCER and CAUSE.
  The subject tends to be definite.
  The subject tends to be topical.
  ...

- **Syntax: “Coding” properties**
  The subject may be formally marked by position and/or morphology (case, agreement).
  ...

- **Syntax: “Behavioural” properties**
  The subject is more prominent than other arguments of the verb. Its prominence is manifested in a variety of phenomena, e.g. reflexive binding.
  Subjects participate in many syntactic processes such as raising, control, relativization, passivization ...

⇒ Subject is a privileged syntactic argument.

- In transformational theories subject is defined structurally (configurationally): it is associated with a certain (high) position on a tree and this conditions its syntactic behaviour.

Some properties of the subject explained by this position:
  - structural prominence expressed by the notion of c-commanding (hence reflexive binding)
  - this position is always projected whether or not a semantic role is assigned to it (expletive subjects)
  - this is the only argument position which may function as a potential target for movement (passive etc.)
  - subject-verb agreement is “checked” in this position

- In non-transformational theories subjecthood is a syntactic primitive. Language-particular diagnostics identifying the subject appeals to behavioural grammatical properties, but not necessarily the position. It varies considerably across languages, and not all relevant properties may be operational in a particular language.

- There is no one-to-one correspondence between morphological encoding and syntactic properties.

2. The morphological encoding of core participants

- **Macro-roles (proto-roles)**

  - Languages often lump semantic roles into three fundamental roles which are assumed in order to explain the different marking systems: S, A, P (or O)
    - S = the single argument of a canonical intransitive verb
    - A = the agent-like argument of a canonical transitive verb
    - P = the patient-like argument of a canonical transitive verb

  - Note that there is some discrepancy in the use of P and O:
    - P used by Comrie (1978) and much of subsequent European tradition
    - O used by Dixon (1979) and much of subsequent Australian tradition
• Attested alignment systems (system of the morphological encoding of participant roles)


One further logical possibility, where A and P are marked in the same way in opposition to S, is extremely rare and is always a partial system. AP/S systems are attested as transitional states from ergative-absolutive to nominative-accusative in some Iranian languages (see Song 2001: 146 for further details and references).

• NB: head vs. dependent marking (in the terminology of Nichols 1986)

- Nominative-accusative (dependent marking): Huánuco Quechua (Quechuan; Peru) (Payne 1997: 134)

(1) a. Juan-Ø Pedro-ta maqan b. Juan-Ø aywan
Juan-NOM Pedro-ACC hits Juan-NOM goes
A P S
‘Juan hits Pedro.’ ‘Juan goes.’

Nominative is usually Ø or relatively little marked, while accusative always has some positive form of marking e.g. suffix.

- Nominative-accusative (head marking): Swahili (Bantu; Tanzania) (Vitale 1981)

(2) a. Ahmed a-li-m-piga Badru b. Fatuma a-li-aguka
Ahmed he-PST-him-hit Badru Fatuma she-PST-fall
A P S
‘Ahmed hit Badru.’ ‘Fatuma fell.’

- Ergative-absolutive (dependent-marking): Diyari (Pama-Nyungan; Australia) (Austin 1981)

(3) a. Karna wapayi b. Karna-li wilha nhayiyi
man goes man-ERG woman sees
S A P
‘The man is going’ ‘The man sees the woman’

Absolutive is usually Ø or relatively little marked, while ergative always has some positive form of marking e.g. suffix.

- Ergative-absolutive (head-marking): Oluta Popoluca (Mixe-Zoque; Mexico)

this 3ERG=eat-INCOM only meat rotten
A P
‘This one eats only rotten meat’

b. weka Ø=pitzüm-pa=k lime:ta-pi
frog 3ABS=exit-INCOM=ANIM bottle-LOC
S
‘The frog is coming out of the bottle’

- Although rare, languages which have both head and dependent marking on an ergative-absolutive system do exist. For an example and references see Song (2001: 145).

• Split-ergative systems
- **Tense-based split-ergative systems**: if there is a TAM-based split we find NOM-ACC or neutral in the present/imperfective and ERG-ABS in the past/perfective, as in Georgian (Kartvelian; Georgia) and several Indo-Iranian languages. See Comrie (1978: 351-2) and Dixon (1994: 100-1) for examples of this phenomenon.

- **NP-based split-ergative systems**: typically reflect Silverstein’s (1976) hierarchy (animacy hierarchy)

  1\&2 prons > 3 prons > proper names & kin terms > human CNs > animate CNs > inanimate CNs

  NOMINATIVE-ACCUSATIVE > NOMINATIVE-ACCUSATIVE < ERGATIVE-ABSOLUTIVE

- Languages make split at different points on the hierarchy, but opposite situations to the hierarchy are not attested.

- For instance, Dyirbal has the ergative-absolutive pattern for 3\textsuperscript{rd} person NPs and nominative-accusative pattern for 1\textsuperscript{st} and 2\textsuperscript{nd} person pronouns (speech act participants).

(5) a. Ngaja balan jugumbil baninyu.
   S S S V(intrans)
   ‘I, the woman, came.’

   b. Ngayguna balan jugumbil balgan.
   P P P V(trans)
   ‘I, the woman, was hit.’

   c. Ngaja banggun jugumbiru bayi yara balgan.
   A A A P P V(trans)
   ‘I, the woman, hit the man.’

- Possible explanation: inherent topicality and agentivity of the positions left on the hierarchy. They are more likely to be associated with the A role and be encoded by the unmarked case for agents (nominative). In contrast, the positions on the right of the hierarchy are prototypical patients (P) and tend to be uncoded by the unmarked patient case (absolutive).

- There are languages (e.g. some Austronesian languages) which demonstrate ergative-absolutive pattern with respect to case marking, but a nominative-accusative pattern with respect to agreement.

  - See Song (2001: 149-50) for discussion of some further types of split.

- **Tripartite systems** (dependent marking): Wangkumara (Pama-Nyungan; Australia) (Song 2001)

(6) a. kana-u[u] kalka-ŋa tiiti-nini
   man-ERG hit-PST dog-\textsuperscript{ACC}
   ‘The man hit the bitch.’

   b. kana-ia palu-ŋa
   man-NOM die-PST
   ‘The man died.’

- **Split-S system**

  - Splits may be determined by the semantic type of the verb (split-S in the strict sense). This type of system is also known as split-intransitive, active-inactive, active-stative, active-static, active-neutral, stative-active, agentive-patientive etc. (Mithun 1991: 551).

  - Splits may exhibit some flexibility depending on agentivity or volitionality of participants, i.e. controlled activity vs. state/uncontrolled activity (fluid-S system). Other factors can also play a role, e.g. animacy, definiteness, topicality, tense/aspect/mood of clause. This is sometimes called Differential Subject Marking.
- Strict split-S systems are more common than fluid split-S systems (Mithun 1991: 542).

● Examples: strict split-S. Head marking in active-inactive systems is more common than dependent marking (Nichols 1992, Siewierska 1996).

- Active-inactive, split-S (head-marking): Mohawk (Iroquoian; Canada) (Mithun 1991: 529-531)

(7) a. transitives
   ik-shereʔs  ‘I chase it’ (1SG A)
   wáʔk-shereʔs  ‘It chases me’ (1SG P)

b. S_p intransitives
   wák-atshenu:ni  ‘I’m happy’
   wák-őreʔsaʔ?  ‘I’m fat’
   wák-lňyo:tskareʔ?  ‘I have hiccups’

c. S_A intransitives
   k-enákereʔ?  ‘I reside (somewhere)’
   k-atáwas  ‘I swim’
   k-atá:tis  ‘I talk’

- Active-inactive, split-S (dependent marking): Laz (Kartvelian; Turkey) (Song 2001: 151)

(8) a. bere-k  imgars  3SG.cry
   S_A  child-ERG
   ‘The child cries.’

b. bere-Ø  oxori-s  doskidu  3SG.stay
   S_p  child-NOM house-DAT
   ‘The child stayed in the house.’

c. baba-k  mečaps  skiri-s  cxeni-Ø
   3SG.give.3SG.3SG
   A  father-ERG child.DAT
   P  horse-NOM
   ‘The father gives a horse to his child.’

● Examples: Fluid splits conditioned by volitionality (dependent marking): Batsbi (North Caucasian; Georgia) (Comrie 1978: 366-7)

(9) a. tyo  naizdrax  qitra
   we.ABS  to-the-ground  fell
   S_p  ‘We fell to the ground (unintentionally, not our fault).’

b. atyo  naizdrax  qitra
   we.ERG  to-the-ground  fell
   S_A  ‘We fell to the ground (intentionally, through our own carelessness).’

● Inverse alignment

- Case-marking or agreement does not indicate the semantic role of the participant. Instead the participant higher on animacy hierarchy automatically claims certain type of marking.

- Direct alignment is a default: in this case the highest participant on the hierarchy is A, while the lower participant is P.

- Inverse alignment applies when the participant lower on the hierarchy is A and the higher participant is P. This is marked on the verb.

- Inverse systems appear to be restricted to head marking only (Nichols 1992; Siewierska 1996).
Example: Plains Cree (Algonquian; Canada) (Foley & van Valin 1985: 297-9)

(10) a. ki-tasam-i-n
    2-feed-DIR-1
    A     P   A
     ‘You feed me.’

   b. ki-tasam-iti-n
    2-feed-INV-1
    P   A
     ‘I feed you.’

● Some languages also make a distinction between proximate (higher empathy/topicality) and obviative (lower empathy/topicality) third person.

Obviative is sometimes referred to as ‘the fourth person’. When two third persons interact, one must be proximate and the other obviative.

If the more topical proximate one acts on the obviative, the direct verb form is used (11a); if an obviative one acts on a proximate one, the inverse form is used (11b). Note that in (11b) the obviative noun is A, but we only know this is by the use of the inverse on the verb.

(11) Plains Cree (Algonquian; Canada) (Foley & Van Valin 1985: 298)

a. asam-e:-w napew-Ø atim-wa
    feed-DR-3 man-PROX dog-OBV
     ‘The man feeds the dog.’

   b. asam-ekw-w napew-wa atim-Ø
    feed-INV-3 man-OBV dog-PROX
     ‘The man feeds the dog.’

3. Behavioural properties

● In many languages the subject is the argument which agrees with the verb, or is required as the antecedent of a reflexive pronoun, or is absent in a subordinate clause under identity with a matrix clause argument (this is called ‘the relation of control’), or is absent under identity with an argument of a coordinated clause.

● Binding relations in Hindi are constrained by the grammatical function of the antecedent. The antecedent of a non-reflexive pronoun cannot be the subject of its clause:

(12) Vijay-ne Ravi-ko uskii saikil par bit.haayaa
    Vijay-Erg Ravi-Obj his bicycle Loc sit.Caus.Perf
     ‘Vijay, seated Ravi on his* bike. (Mohanan 1994: 126)

● In Japanese some verbs take special forms to honour the subject referent, and not other grammatical functions. The verb form o-V ni naru is used to honour the subject sensei ‘teacher’, and cannot be used to honour a nonsubject.

(13) a. sensei wa hon o-o-yomi ni narimashita
    teacher Top book Acc Honorific-read Cop become.Polite.Past
     ‘The teacher read a book.’

   b. *Jon wa sensei no o-tasukere ni natta
    John Top teacher by Honorific-help Cop become.Past
     ‘John was saved by the teacher.’ (Matsumoto 1996: 27-28)

● Some subject tests in English:

(14) Control structures

Fred wants [ ____ to come]
Fred wants [ ____ to see Henrik]
*Fred wants [ Henrik to see ____ ]
(15) Clause coordination

Fred went to the store and [ ____ came back]
Fred went to the store and [ ____ bought a hamburger]
*Fred went to the store and [Henrik saw ____ ]

(16) Secondary predicates

Fred saw Mary [ ____ swimming]
Fred saw Mary [ ____ eating cheese]
*Fred saw Mary [Henrik hitting _____ ]

● Diversion: taxonomy of control and raising constructions in English

subject-to-subject raising: John, is likely [Ø, to leave]
subject-to-object raising: I want John, [Ø, to leave]
object-to-subject raising: John, is easy [to please Ø,]
subject control: John, is reluctant [Ø, to leave]
object control: John persuaded Bill, [Ø, to leave]

No clear examples of structures where the controlled element is not the subject of the embedded clause.

The major difference between control and raising predicates is in semantics and argument structure:

(17) a. It is likely [that John leaves]
   proposition
   experincer
   action

b. John is reluctant [that John leave]

● None of these properties is definitional for subjecthood, and indeed all of these properties have been found to be associated with nonsubjects as well as subjects, depending on the language. Only the whole cluster of properties unambiguously identifies subject.

● NB: Behavioural and coding properties do not necessarily come together. In terms of the morphological encoding the subject is not necessarily nominative. These are “quirky case” subjects, e.g. in Icelandic.

(18) a. Drengina vantar mat
     boys.ACC lacks food.ACC
     ‘The boys lack food.’

b. Henni hefur þótt Ólafur leiðinlegur
   she.DAT has thought Olaf.NOM boring
   ‘She has considered Olaf boring.’

(19) Subjecthood tests: reflexive binding

Henni þykir bróðir sinni/*hennar leiðinlegur
she.DAT thinks brother self’s/*her boring
She considers her (self’s) brother boring.

(20) Subjecthood tests: subject-to-subject raising (the case of the lower subject remains the same)

a. Hann virðist vanta peninga
   she.ACC seems to.lack money.ACC
   ‘She seems to lack money.

b. Barninu virðist hafa batnað veikin
   child.DAT seems to.have recovered.from disease.NOM
   ‘The child seems to have recovered from the disease.’
4. Subjects and topichood

- There is a strong association between topics and subjeckhood (Keenan 1976; Givón 1976; Comrie 1989; Erteschik-Shir 2007, among others). This reflects a tendency towards isomorphism between pragmatic (informational), semantic, and syntactic prominence.

- Lambrecht (1994:132–137) argues that subjects are unmarked topics. Non-topical subjects must be structurally marked. That is, the subject referent is interpreted as topic unless there are syntactic or prosodic cues to the contrary, such as appearance of a topical nonsubject element in a noncanonical sentence-initial position, or prosodic prominence of focus subjects.

(21) a. How’s your neck? My neck/It HURTS.
    b. What’s the matter? My NECK hurts. (Lambrecht 1994:137)

- Some languages place stricter restrictions on the relation between syntax and information structure: they disallow nontopical subjects altogether. In Dzamba (Bantu) all nonsubject arguments can be questioned in situ, but the subject cannot be questioned in its initial position: it must be questioned using a relative clause. Since wh-questions are inherently non-topical, this indicates that Dzamba does not allow non-topical subjects, and indeed Bresnan & Mchombo (1987) claim that Dzamba subjects are grammaticalised topics and for this reason cannot be questioned.

- In some languages, a nontopical or focused argument may not be realised as subject, but must be demoted to a lower function on the hierarchy of grammatical functions. In constructions where it is not topical, the thematically highest argument can be associated with the behavioural features of objects. For instance, Creissels (2008: 157) shows that in French impersonal constructions, the argument of an intransitive verb can appear postverbally. He refers to such cases as “pragmatic conditioning on fluid transitivity”.

(22) a. une femme viendra
    b. il viendra une femme
    ‘A woman will come.’
    ‘A woman will come (there will be a woman coming).’

- In Northern Khanty (Uralic) the transitive subject must bear the topic role. When the A argument is nontopical, passivisation is required and it become a locative-marked oblique.

(23) a. What about Peter?
    b. (luw) Juwan-na re:sk-ş-s-a
       he John-Loc hit-Past-Pas.3SgSubj
       ‘John hit him.’ (literally: ‘He was hit by John.’)
    c. (luw) Juwan-na ke:si-na ma-s-a
       he John-Loc knife-Loc give-Past-Pas.3SgSubj
       ‘John gave him a knife.’ (literally: ‘He was given a knife by John.’)

(24) a. kalaŋ xoj-na we:l-s-a?
    reindeer who-Loc kill-Past-Pas.3SgSubj
    ‘Who killed the reindeer?’
    b. Juwan-na we:l-s-a
       John-Loc kill-Past-Pas.3SgSubj
       ‘John did.’

- Fluid splits in morphological encoding can be conditioned by topicality.

In some languages, subjects which are also topics are grammatically marked; nontopical or focus subjects are unmarked. This is Differential Subject Marking which depends on topicality.
Topical subjects often receive special casemarking, and also tend to trigger more agreement than nontopical or focused subjects (Comrie 2003, Siewierska 2004:159ff, Corbett 2006:197-204).

● In Somali (Semitic) there are two subject forms, the ‘absolutive’ case and nominative case (Saeed 1984, 1987). The absolutive is the basic citation form and marks some subjects, objects and obliques. The nominative is derived from the absolutive by tonal alternations and sometimes by adding a final vowel. The nominative is found only on topical subjects, while nontopical or focused subjects must stand in the absolutive form.

With focused subjects the verb must be in the so-called ‘relative’ form, a paradigm in which agreement is reduced: it includes only three distinct forms, in contrast to the five or six forms typically found in paradigms used with topical subjects.

(25) a. nínku wuu imánayaa
    man.Nom Decl.3Sg.Masc come.Prog.3SgMasc
    ‘The man is coming.’

b. nínka ayàa imánayá
    man.Abs Foc come.Rel.1Sg/2Sg/3SgMasc/2Pl/3Pl
    ‘The MAN is coming.’

● Subject is the most prominent argument of the verb and is the most active in syntax. It tends to be prominent is semantics and information structure too, but languages exploit this relationship in different ways.

Further references