APPRAOCHES TO
GRAMMATICALIZATION

Volume II

Edited by

ELIZABETH CLOSS TRAUGOTT
and BERND HEINE

Offprint
The Role of Motivation in the Emergence of Grammatical Categories: the Grammaticization of Subjects

Marianne Mithun

University of California, Santa Barbara

INTRODUCTION

Our search for forces motivating the grammaticization of particular categories has uncovered two main kinds of factors: Cognitive and communicative. Speakers tend to develop linguistic structures that mirror their cognitive structures in order to have efficient tools for the expression of thought. The ongoing activity of communication hones these structures: Patterns that are used frequently become routinized and automatic. If all human beings have similar cognitive structures and communicative needs, we might expect the kinds of categories grammaticized in different languages to be the same. Certain grammatical categories do appear in language after language, yet one of the most intriguing aspects of language is the fact that grammatical systems are far from universal.

Understanding the processes of grammaticization involves not only discovering which categories tend to be grammaticized in languages, but also why these are not grammaticized in every language. Cognitive structures may set up predispositions for the grammaticization of certain categories rather than others, but they are not mandates. Whether a preferred category is grammaticized or not can be affected by several factors. First, it may depend simply on chance. Second, it may depend on cultural preoccupations. For Karok speakers, for example, the basis of spatial orientation was apparently the Klamath River. Reference to the river must have appeared sufficiently often in everyday speech to become routinized, ultimately yielding a rich system of locative suffixes with such meanings as 'hence upriverward', 'hither from upriver', 'hence across a body of water', 'hither from across a body of
water’, ‘hence downriverward’, ‘hither from downriver’, etc. (Bright, 1957: 95). In many other cultures, orientation in terms of a river is so rare that it will probably always remain expressible only lexically. Third, whether or not a category is grammaticized at a particular moment must depend upon the shape of the current grammar. The formation of new grammatical categories is motivated or hindered by the contours of the existing grammatical system.

As is well known, it was once proposed that all languages are “deeply accusative”: Whatever the organization of their morphological case marking, their syntax is always accusatively based (Anderson, 1976). Some grammatical theories are still based on this notion. Of course the basic grammatical category in accusative systems is that of subject. In accusative systems, if any case is formally unmarked, it is normally the subject. If any case is required in all clauses, it is the subject. The citation form of nouns is usually the subject form. The function of subjects is clear: They are essentially grammaticized clause topics. Accusative case organization is accordingly prevalent in constructions involving clauses and combinations of clauses: Syntactic structures. In fact, many languages with alternative case categories in their morphologies do exhibit accusatively organized syntax. The formal grammaticization of a subject category is not universal among languages, however. Some languages exhibit no evidence of an accusative base, even in their syntactic patterns, as has been demonstrated by Li and Thompson (1976), Dixon (1979), and Bossong (1984) among others.

The grammaticization of a subject category is not always a matter of chance. Whether or not a particular language develops this category can be affected by structural features elsewhere in the grammar. One of these is the presence of a full set of obligatory, morphologically bound pronouns, in at least three persons and two cases, based on case distinctions other than nominative versus accusative. In what follows, the case systems of two different types of non-accusative languages will be examined: One agent/patient and one ergative/absolutive. Neither has formally grammaticized a subject category within either the morphology or syntax. It will be shown that this lack of grammaticization is related to common characteristics of their grammars that diminish the motivation for such a development.

1. CODING

The grammaticization of a case system in a language is most obvious in the coding, i.e. formal marking of case categories. The case suffixes on nouns
in Latin and Russian, for example, provide clear evidence of a subject category in those languages. The forms of their pronouns provide similar evidence: Pronouns referring to subjects have different forms from those referring to objects, possessors, etc.

In Cayuga, an Iroquoian language, formal case marking is concentrated in the pronominal system. All verbs contain pronominal prefixes referring to their core or primary arguments. The forms of the prefixes reflect an agent/patient or active case system. Intransitive verbs contain a single pronominal prefix whose form reflects the semantic role of the single core participant. When an event involves only an agent, the pronominal prefixes have the forms in (1). The data cited here come from the speech of Reginald Henry of Six Nations, Ontario.

(1) Cayuga intransitives: Agents

\[ \text{kqhté-ke}\, 'I am leaving' \quad \text{katkéhehs}\, 'I get up' \]
\[ \text{sqhté-ke}\, 'you are leaving' \quad \text{satkéhehs}\, 'you get up' \]
\[ \text{hahté-ke}\, 'he is leaving' \quad \text{hatkéhehs}\, 'he gets up' \]

When the event involves only a patient, a different set of pronominal prefixes is used. Compare the pronouns in (2).

(2) Cayuga intransitives: Patients

\[ \text{akáht?agh}\, 'I have gotten full' \quad \text{akétshahni?k}\, 'I am afraid' \]
\[ \text{sqhtá?qh}\, 'you have gotten full' \quad \text{satsháhni?k}\, 'you are afraid' \]
\[ \text{hohtá?qh}\, 'he has gotten full' \quad \text{hotsháhni?k}\, 'he is afraid' \]

Transitive verbs in Cayuga contain complex pronominal prefixes referring to both agent and patient. (Agent and patient components of most of the transitive prefixes have become fused morphologically.)

(3) Cayuga transitives: Agents and patients

\[ \text{aské-ke}\, 'you saw me' \quad \text{aha\cdot ke-ké}\, 'he saw me' \]
\[ \text{ahéhsé-ke}\, 'you saw him' \quad \text{ahé-ke}\, 'I saw him' \]
\[ \text{ashé-ke}\, 'you saw her' \quad \text{asha\cdot kó-ke}\, 'he saw her' \]

(Cayuga contains many more prefixes than those cited here, around sixty in all. Pronouns distinguish singular, dual, and plural number, first, second, and third person, exclusive and inclusive in first person, and masculine, feminine, and neuter gender in third person, in addition to agent and patient case.) The classification of primary participants into agents and patients is semantically based, but it is fully grammaticized. Some agents may seem semantically more
active than others: A person who cuts a rope seems more agentive than one who sees a rope, but they are categorized equivalently as agents by the grammar. A rope that is cut seems more affected than one that is only seen, but both are classified as patients grammatically.

These prefixes are not simple agreement markers: They are full referential pronouns, representing the core arguments themselves. Verbs containing them constitute grammatically complete clauses on their own. When separate noun phrases are also present in a clause, the pronominal prefixes remain in place. The noun phrases function essentially as appositives to the pronouns and bear no case marking (Jelinek, 1984; Mithun, 1986a; 1986b).

(4) Cayuga clauses with appositive noun phrases

\[ \text{Ha-hni ne? so'wa's} \]
\[ \text{MASC.AGT-bark.HABITUAL the dog} \]
\[ \text{he-bark-s the dog} \]

'The dog is barking.'

\[ \text{Hy-?ani a-he-ke-?} \]
\[ 3/2-father AOR-1/MASC-see-PUNCTUAL \]
\[ \text{your father I saw him} \]

'I saw your father.'

Selayarese, an Austronesian language of Indonesia, also shows no subject category in its formal marking. Like Cayuga, it contains obligatory bound pronouns, but these operate on an ergative/absolutive basis. The Selayarese data cited here come from the speech of Hasan Basri of South Sulawesi.

The single argument of an intransitive verb is represented by an absolutive pronominal enclitic. Note that the forms of the absolutive enclitics are the same whether they refer to agents or patients.

(5) Selayarese intransitives: Absolutive enclitics

\[ a?liumpa?a \quad \text{‘I jump’} \]
\[ a?liumpakko \quad \text{‘you jump’} \]
\[ a?liumpakki \quad \text{‘we (incl) jump’} \]
\[ a?liumpakkay \quad \text{‘we (excl) jump’} \]
\[ a?liumpai \quad \text{‘he/she/they jump’} \]
\[ ma?nya\text{a} \quad \text{‘I am tired’} \]
\[ ma?nya\text{ko} \quad \text{‘you are tired’} \]
\[ ma?nya\text{ki} \quad \text{‘we (incl) are tired’} \]
\[ ma?nya\text{kan} \quad \text{‘we (excl) are tired’} \]
\[ ma?nya\text{in} \quad \text{‘he/she/they are tired’} \]
In transitive verbs, the absolutive enclitic refers to the semantic patient. The semantic agent is indicated by an ergative prefix.

(6) Selayarese ergative prefixes and absolutive enclitics

<table>
<thead>
<tr>
<th>Selayarese</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>kuisse'1</td>
<td>'I know him'</td>
</tr>
<tr>
<td>muisse'1</td>
<td>'you know him'</td>
</tr>
<tr>
<td>riisse'1</td>
<td>'we (incl) know him'</td>
</tr>
<tr>
<td>toisse'1</td>
<td>'we (excl) know him'</td>
</tr>
<tr>
<td>laisse'1</td>
<td>'he/she/they know(s)'</td>
</tr>
</tbody>
</table>

(Neither number nor gender is distinguished in Selayarese pronouns. The inclusive form also functions as a formal second person pronoun. Verbs unmarked for tense are used to describe both present and past events.)

In Selayarese, as in Cayuga, verbs with bound pronouns constitute grammatically complete clauses in themselves. When separate noun phrases are also present in the clause, the bound pronouns remain. The noun phrases are unmarked for case.

(7) Selayarese clauses with appositive noun phrases

<table>
<thead>
<tr>
<th>Selayarese</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mây'yay-i hásay aló-nni</td>
<td>tired-3.ABS Hasan day-this</td>
</tr>
<tr>
<td></td>
<td>he is tired Hasan today</td>
</tr>
<tr>
<td></td>
<td>'Hasan is tired today.'</td>
</tr>
<tr>
<td>La-pál'lu-i berasá-ño i-hásay</td>
<td>3.ERG-cook-3.ABS rice-the Mr-Hasan</td>
</tr>
<tr>
<td></td>
<td>he cook it the rice Mr. Hasan</td>
</tr>
<tr>
<td></td>
<td>'Hasan cooked the rice.'</td>
</tr>
</tbody>
</table>

Neither Cayuga nor Selayarese thus exhibits any evidence of a subject category in its formal case marking. Morphological case marking is agent/patient based in Cayuga and ergative/absolutive based in Selayarese.

2. BEHAVIOR

Evidence of the grammaticization of a category need not be limited to formal markers. If certain kinds of constituents all behave grammatically in a particular way or if they trigger certain grammatical patterns, and no others do, this group is often said to constitute a grammatical category. The hypothesis that all languages are "deeply" or syntactically nominative/accusative was proposed on this basis.
The most common syntactic areas in which evidence of the grammaticization of subject has been sought are in imperatives, coordination, and subordination.

2.1. Commands

Imperative constructions have been taken as evidence of the grammaticization of a subject category in many languages. It has been suggested that in English, for example, subjects pattern as a group in that they are systematically deleted from imperatives. Second person subject pronouns are omitted from all commands, whether intransitive (___ eat!) or transitive (___ eat your beans!).

Cayuga offers no equivalent evidence of a grammaticized subject category, because pronominal prefixes are present in every verb, including imperatives. Note the second person pronouns in the commands in (8).

(8) Cayuga imperatives with bound pronouns

\[ \text{San-ahd-owe.k, s-ak-ya?t-awi?-t hni?} \]
\[ \text{2.AGT-SEMI.REFL-crown-cover 2.AGT-body-encircled-CAUS too} \]
\[ \text{you put on your hat} \quad \text{you put on your coat too} \]
\[ \text{‘Put on your hat and coat.’} \]

Selayarese formal imperatives are structurally equivalent to those in Cayuga. In formal commands, the bound pronouns remain, whether the commands are transitive or intransitive.

(9) Selayarese formal imperatives with bound pronouns

\[ \text{Máne-ko} \]
\[ \text{go-2.ABS} \]
\[ \text{you-go} \]
\[ \text{‘Please go.’} \]
\[ \text{Bérasa mu-pállu-i} \]
\[ \text{rice 2.ERG-(TRANS).cook-3.ABS} \]
\[ \text{rice \text{you cook it}} \]
\[ \text{‘Please cook the rice.’} \]

Informal commands, however, may appear without pronouns. Absolutive pronouns may be omitted from intransitive verbs, and ergative pronouns from transitive verbs.
MOTIVATION IN THE EMERGENCE OF GRAMMATICAL CATEGORIES

(10) Selayarese informal commands without pronouns

Máge(-θ)
(θ)go
'Go!'

(θ-)Pállu-i
(θ)-(TRANS).cook-3.ABS
'Cook it!'

These informal commands might at first be taken as behavioral evidence of a subject category in Selayarese. The category represented here is not subjecthood, however, but agency or control, as pointed out by Dixon (1979), Anderson (1979), and others. Just as in English, commands can only be felicitously addressed to someone in a position to carry them out. When the second person is not in control, Selayarese verbs with second person pronouns constitute statements, not formal commands. If the pronouns are omitted, the result is not an ungrammatical informal command, but a simple verb stem, pragmatic nonsense.

(11) Selayarese pragmatic pattern: Control

máŋŋąko ‘you are tired’ máŋŋąŋ ‘to be tired’
bássoro ko ‘you are full’ bássoro ‘to be full’

Selayarese commands provide no more evidence of a grammaticized subject category than their English counterparts. In both languages, second persons in control, not subjects, are omitted from informal commands. This represents a pragmatic pattern rather than a syntactic restriction.

2.2. Coordination

A second area in which behavioral evidence of a particular case organization is sometimes sought is clause coordination. Dixon (1972) proposed that in Dyirbal, for example, clauses sharing a coreferent nominal can be conjoined only if that nominal appears in absolutive case in both clauses. This restriction was considered evidence of a grammaticized absolutive category.

The omission of coreferent nominals from coordinate constructions has also been cited as evidence of certain case categories in various languages. In English conjoined sentences, for example, the subject of the second clause may be omitted if it is coreferent with the subject of the first. Speakers can thus say either of the two sentences below.

(12) a. He picked up his suitcase and he left
b. He picked up his suitcase and left
This pattern is restricted to subjects. English speakers do not omit the second of two direct objects.

(13) *He picked up his suitcase and carried out

The noun phrase omitted from the second clause must be coreferent with the subject of the first clause, not some other constituent. There is no doubt from the sentence below about who left.

(14) He kissed her and left.

Cayuga exhibits no syntactic restrictions on the argument structure of coordinate clauses. Any pragmatically relevant clauses may be conjoined. Furthermore, since all Cayuga verbs contain pronominal prefixes, there is no deletion of coreferent noun phrases under coordination. The pronominal prefixes in the verbs in (15) refer to the same second person singular agent, but neither can be omitted.

(15) Cayuga conjunction with coreferent agents

\[
\begin{align*}
eg &-t\text{-}hs\text{-}e\text{-}? & ki? & kye\text{-}? & wa\text{?}tsi \\
\text{FUT-CISLOC-2.SG.AGT-go-PUNCTUAL} & \quad \text{just then later} \\
\text{you} & \quad \text{will come} \\
\quad & \text{just then later} \\
\end{align*}
\]

\[
\begin{align*}
eg &-t\text{-}hs\text{-}haw\text{-}i\text{-}ht\text{-}ahk. \\
\text{FUT-CISLOC-2.SG.AGT-carry-CAUS-INST-PUNCTUAL} \\
\text{you} & \quad \text{will carry with you} \\
' & \text{Why don’t you just come on over later, then, and bring it along.’}
\end{align*}
\]

Selayarese, like Cayuga, exhibits no syntactic restrictions on the argument structure of coordinate clauses. Bound pronouns are also not omitted under conjunction. In (16) for example, pronouns remain with each clause, even though they are coreferent.

(16) Selayarese conjunction

\[
\begin{align*}
M &-ge\text{-}i & ku\text{-}alle & an\acute{a}\text{k}-ku, & nampa\text{-}a & m &uliam & m &ge \\
go\text{-}3.ABS & 1.ERG\text{-}take & child\text{-}1.POSS & then\text{-}1.ABS & go & toward \\
go her & I & take & my daughter & then & I & go & toward \\
ri & s\acute{a}po & nampa\text{-}a & tinro. \\
to house & then\text{-}1.ABS & sleep \\
to house & then & I & sleep \\
'I & picked & up & my daughter, & went & home, & and & went & to & sleep.’
\end{align*}
\]

All would be subjects in English, as can be seen from the translation.
Since neither Cayuga nor Selayarese contains syntactic restrictions on the argument structure of conjoined clauses, and no pronouns are omitted from conjoined clauses, conjunction provides no evidence of the grammaticization of a subject category in either language.

2.3. Nominalized clauses

In many languages, clauses are subordinated by nominalization. This is often accomplished with possessive forms. In English nominalized clauses, the subject appears as a possessor whether it functions semantically as an agent or patient, and whether the clause is intransitive or transitive.

(17) I left (Agent of intransitive) 
[My leaving] disturbed him

(18) I was sick (Patient of intransitive) 
[My being sick] slowed things down

(19) I threw out the comics (Agent of transitive) 
[My throwing out the comics] was a mistake

Cayuga exhibits no comparable morphological nominalization of clauses. In Selayarese, however, nominalization with possessives is the most common subordinating device.

Basic possession in Selayarese is shown by pronominal suffixes on nouns, as in (20).

(20) Selayarese possessive pronominal suffixes

\[
\begin{align*}
\text{asúŋku} & \quad \text{‘my dog’} \\
\text{asúmmu} & \quad \text{‘your (familiar) dog’} \\
\text{asúnta} & \quad \text{‘our (inclusive) dog’} \\
\text{asúmba’} & \quad \text{‘our (exclusive) dog’} \\
\text{asúńna} & \quad \text{‘his/her/their dog’}
\end{align*}
\]

The same set of possessive suffixes is used in nominalized clauses. When the nominalized clause is intransitive, the possessor is of course the single participant, the absolutive. (The absolutive enclitic appears on the first element of the clause.)
When the clause is transitive, the possessor is the semantic patient, again the absolutive.

(22) Selayarese transitive clause nominalization
*Tàjat-a rinni [sàngèn-na ku-ʔuŋpa ammám-mu].
await-1.ABS at-this [until-3.POSS 1.ERG-find father-2.POSS]
‘Wait here until I find your father (until his finding).’

Subordination by nominalization thus provides no evidence of a grammaticized subject category in either Cayuga or Selayarese. In Cayuga, it yields no evidence of any category at all, and in Selayarese, it is consistently absolutive based.

2.4. Complex constructions

Evidence of a subject category has been sought in some languages in patterns sometimes referred to by “equivalent noun phrase deletion” and “raising”. It has been proposed that in English for example, the subject of a complement clause is omitted if it is coreferent with the subject of its intransitive matrix clause.

(23) Sally hopes [Hilda will win]
(24) Sally hopes [to win]

Subjects of complements are also assumed to be omitted if they are coreferent with the object of a transitive matrix clause. In the sentence below, the performer must be Hilda.

(25) Sally invited Hilda [to perform]

Cayuga exhibits no coreferent deletion at all, since all verbs contain obligatory pronominal prefixes referring to their agents and/or patients. All of the first person agent pronouns must remain with their verbs in the sentence in (26), even though they would be coreferent subjects in English.
(26) Cayuga complement of intransitive

\[ A'yè^? k-áhskáné-s o-thowêe-ké \]

**INFER** 1.AGT-want-HAB N-cold-LOC

seems I wish at it’s cold

\[ h-ák-e-? \]

**TRANSLOC-OPT-1.AGT-go-PUNC**

I would go there

\[ a-k-ê-nat-áê-hna-? \]

**OPT-1.AGT-SEMI.REFL-camp-lay-PURP-PUNC**

‘I want to go up north I would go camp to camp.’

All pronominal prefixes also remain in place when the matrix clause is transitive.

(27) Cayuga complement of transitive

\[ e-kashe-hókâ-o? hni? \]

**FUT-2/3.PL-invitation-give-PUNC** too

you will invite **them** too

\[ e-t-kap-te-kh-ôny-âhne-? hwis-hó?ph. \]

**FUT-CISLOC-3.PL-REFL-meal-make-PURP-PUNC** five-DISTR

**they** will come eat five-ish

‘Invite them to come eat around five o’clock, too.’

Selayarese differs in an interesting way from Cayuga. When both the matrix and the complement clause are intransitive, the complement verb may contain no clitic, as in (28).

(28) Selayarese intransitive complement

\[ A?jáñi-a la-minâhanj. \]

**promise-1.ABS FUT-(INTR).follow**

I promise will follow

‘I promise to follow.’

This might at first appear to exemplify coreferent subject deletion. When the coreferent noun phrase is the agent of a transitive complement clause (ergative), however, no omission is possible.
Selayarese transitive complement
\[ A\text{-}jáñji\text{-}a \quad la-ku-pínáhñ-ko \]
\[ \text{promise-1.ABS FUT-1.ERG-follow-2.ABS} \]
\[ \text{I promise I will follow you} \]
\[ \text{‘I promise to follow you.’} \]

Only coreferent absolutes can be omitted.

A second kind of subordinate construction, sometimes called “raising”, has also been cited as evidence of a subject category in some languages. According to raising analyses, underlying subjects of complement clauses appear as derived subjects or objects of matrix clauses.

\[ 30 \]  \text{It [John is happy] seems $\rightarrow$ John seems happy}  
\[ 31 \]  \text{I saw it [that she left] $\rightarrow$ I saw her leave}  

(Underlying subjects of subject complements appear as derived matrix subjects. Underlying subjects of object complements appear as derived matrix objects.)

As usual, Cayuga agent and patient pronouns remain with their verbs, regardless of syntactic context.

\[ 32 \]  Cayuga complement of intransitive
\[ T\text{-}ka\text{-}kó\text{-}t \quad kye\text{-}\emptyset \quad hne\text{-}\emptyset \quad o\emptyset \quad e\text{-}hs\text{-}nín\emptyset\text{-}\emptyset \]
\[ \text{CISLOC-N-necessary INFER CONTR now FUT-2.AGT-buy-PUNC} \]
\[ \text{it is necessary guess but now you will buy} \]
\[ \text{‘But you’ll have to buy that (the lard).’} \]

\[ 33 \]  Cayuga complement of transitive
\[ A\text{-}kó\text{-}ké\text{-}\emptyset \quad s\emptyset\text{-}të\emptyset \quad s\text{-}yethw\text{-}áhsp\text{-}h \]
\[ \text{AOR-1/2-see-PUNC last night 2.AGT-plant-DISTR-STATIVE} \]
\[ \text{I saw you last night you were planting things} \]
\[ \text{‘I saw you planting last night.’} \]

Selayarese presents a slightly different picture. Constructions like the negative sentence below have been analyzed as the result of raising in some related languages. It has been posited that the pronoun ‘I’ originates with the second verb ‘tired’, then is raised to the first, ‘not’.

\[ 34 \]  Selayarese complex predicate
\[ Géle\text{-}a \quad mánnyañ \]
\[ \text{not-1.ABS tired} \]
\[ \text{I not tired} \]
\[ \text{‘I am not tired.’} \]
Similar constructions can be found with other intransitive verbs in Selayarese.

(35) Selayarese complex predicates

*A?ra?-a n-ânre*
want-1.ABS INTR-eat
I want eat
'I want to eat.'

*Máne-i n-rio.*
go-3.ABS INTR-bathe
he go bathe
'He went to take a bath.'

The first verb may be transitive.

(36) Selayarese transitive matrix

*La-jâñjañ-ki r-bisâra*
3.ERG-see-1.INCL.ABS INTR-talk
he saw us talk
'He saw us talk.'

When the second verb is transitive, however, the "raised" pronoun refers to the semantic patient. The construction involves absolutes, not subjects.

(37) Selayarese transitive complements

*Géle-a la-jâñjañ.*
not-1.ABS 3.ERG-see
not-me he see
'He didn’t see me.'

*A?rak-ko ku-jâñjañ.*
want-2.ABS 1.ERG-see
want-you I see
'I want to see you.'

*Mâim-mo-ko ku-saré-âñ.*
finish-PERF-2.ABS 1.ERG-give-DAT
have-you I give it
'I have already given it to you.'

The two patterns described here could be interpreted as absolutive-based syntactic processes, whereby coreferent absolutive arguments of complements are deleted and raised respectively. Absolutive pronouns in Selayarese are
actually enclitics rather than suffixes. They are suffixed to the first element of clauses, not simply to verbs. The first word is often a verb, but not always; note the absolutive enclitic on the conjunction *nampa* ‘and then’ in (16) and on *sangen* ‘until’ in (21). It is not surprising that they should have scope over the entire complex clause.

### 2.5. Relativization

The formation of relative clauses in many languages is sensitive to case, often to the category of subject. In some languages, for example, heads of relative clauses must be their subjects. Neither Cayuga nor Selayarese exhibits any such restrictions.

In Cayuga, strict formal identity between the head and constituents of the relative clause is not necessary. In (38), ‘someone’ would normally be considered third person feminine-indefinite singular, but the associated verb ‘marry’ has a first person plural pronominal prefix.

(38) **Cayuga relativization without formal identity**

\[
\text{The} \ ni? \ i'-\acute{a} \quad a-ke-tshi'-? \quad s\acute{p}k\acute{a}\text{-}?a \\
\text{not just NEG-OPT OPT-1.AGT-find-PUNC someone}
\]

\[
\text{not just impose I would find someone}
\]

\[
a-\text{yokhni-nya}-k.
\]

\[
\text{OPT-1.DU.PAT-marry}
\]

\[
\text{we two would marry}
\]

‘I can’t find anyone to marry me.’

Even when they are equivalent, all pronouns remain with their verbs in both the matrix and relative clauses. (The feminine-indefinite singular is used for a people.)

(39) **Cayuga relativization with formal coreference**

\[
\text{Tho} \ ti? \ ni-\text{-yo}-\text{hi ne-kye}
\]

\[
\text{there just so-it-is this}
\]

‘And that is the way

\[
a-k-\text{yako-ihw-}\acute{a}\text{-et}-a?-s.
\]

\[
\text{AOR-DU-FEM.PAT-matter-lay-INCH-DAT}
\]

\[
\text{one decided}
\]

\[
\text{they decided},
\]
MOTIVATION IN THE EMERGENCE OF GRAMMATICAL CATEGORIES

In Selayarese, an optional prefix to- ‘who’ on the verb of the relative clause indicates that the head is human, nu- that it is non-human. There are no case restrictions on the shared constituent in either the matrix or relative clause. The bag in the sentence below functions as an absolutive in the matrix clause and as a locative in the relative clause.

(40) Selayarese relativization with locative head

\[\text{Ku-}^{\text{na}}\text{ppa-mo-i t\^{a}si (nu)-mu-panaro-i\^{n}jo d\^{o}e?}\]
\[\text{1.ERG-find-PERF-3.ABS bag REL-2.ERG-put-the money}\]
\[\text{I found it bag the which you put money}\]
\[\text{I found the bag in which you put money.}\]

The same bound pronouns appear in relative clauses as in independent clauses, just as in Cayuga.

(41) Selayarese bound pronouns under relativization

\[\text{Ku-}^{\text{na}}\text{ppa-mo-i ana?-\^{a}na?}\]
\[\text{1.ERG-find-PERF-3.ABS child-child}\]
\[\text{I have found him child}\]
\[\text{(to)-la-lukka?-i\^{n}jo-i as\^{u}n-ku}\]
\[\text{that 3.ERG-steal-the-3.ABS dog-1.POSS}\]
\[\text{the who he-stole it my dog}\]
\[\text{I have found the child that stole my dog.}\]

Relativization thus provides no evidence of a grammaticized subject category in either Cayuga or in Selayarese.

3. WHY CAYUGA AND SELAYARESE NEED NOT GRAMMATICIZE A SUBJECT CATEGORY

It is clear why so many languages have grammaticized the notion of clause topic into a category of subject. Topic-comment structures are among the most frequently occurring in natural discourse in any language. Identification of the clause topic is crucial to understanding the point of most speech. The codification of this discourse category into a grammatical category is thus well motivated. Yet neither Cayuga nor Selayarese exhibits evidence of
the grammaticization of a subject category at all, even in syntactic constructions. This is no accident. Certain structures already present in their grammatical systems reduce the motivation for such grammaticization.

As is well known, speakers do not choose clause topics at random. They select one participant as a starting point, then say something about it (Chafe, 1976). This starting point is normally that participant with which the speaker chooses to identify. Speakers generally prefer first person topics to second, and second to third. They prefer humans to nonhumans, and agents to nonagents. Accordingly, the choice of subject, the grammaticized clause topic, usually depends upon person, humanness, and agency.

Most of these distinctions are already obligatorily specified in every clause in Cayuga and Selayarese. Both languages distinguish first, second, and third person within the bound affixes. Cayuga overtly distinguishes not only agents and patients, but humans and nonhumans within pronominal forms. Selayarese distinguishes agents and patients when not otherwise obvious: In transitive clauses. The obligatoriness of just these pronouns insures that the clause topic, the participant that would usually be the subject in an accusative language, is overtly represented in every Cayuga and Selayarese clause.

A subset of core participants in each language is unrepresented by the bound pronouns. Each of these subsets consists of participants that do not normally function as clause topics.

In Cayuga, all agents are represented by pronouns. All human patients are also represented. The transitive pronominal prefix k-he- ‘I/her’ for example, refers to a first person agent (k-) and a feminine patient (-he-).

(42) Cayuga transitive pronoun with human patient

\[
E-k'sa-ʔǎa \quad a-k'he-ʔe-ʔ. \\
\text{FEM.child-DIM AOR-1.AGT/FEM.PAT-see-PUNC}
\]

little girl I saw her

'I saw a/the girl.'

Nonhuman patients are represented except in one context: When an agent is also present. The pronominal prefix k- ‘I/(it)’ contains no overt mention of the inanimate patient. (The following e is epenthetic.)

(43) Cayuga verb with human agent and nonhuman patient

\[
Ka-n̂ghs-ˈt \quad a-k'ɛ-ʔe-ʔ. \\
\text{N-house-stand.STAT AOR-1.AGT-see-PUNC}
\]

house I saw

'I saw a/the house.'

Note that the same prefix is used in intransitive verbs.
MOTIVATION IN THE EMERGENCE OF GRAMMATICAL CATEGORIES

(44) Cayuga verb with human agent only
\[
A\text{-}k\text{-at-awé-hna}^2. \\
AOR\text{-}1.AGT\text{-SEM} \text{.REFL\text{-}bathe\text{-PURP} \text{-PUNC}}
\]
′I went swimming.′

Nonhuman patients are not normally selected as clause topics when agents are also involved, exactly the situation presented by transitive verbs.

In Selayarese transitive clauses, agents are represented by ergative pronouns and patients by absolutive pronouns, whether they are human or not.

(46) Selayarese agent and patient of transitive
\[
Ku\text{-}hálli^2\text{-i } sapó\text{-ñjo}. \\
1.ERG\text{-buy}\text{-3.ABS house-the}
\]
I buy it the house 'I bought the house.'

If the patient of a transitive is indefinite, however, it is not represented by a pronoun. The verb is then formally intransitive, and the agent is absolutive.

(47) Selayarese semantic transitive with indefinite patient
\[
M\text{-}máli\text{-a } sápo. \\
INTR\text{-buy}\text{-1.ABS house}
\]
I buy house 'I bought a house.'

Indefinite patients of semantic transitives are not normally topical. Topics, the starting points of predications, usually represent established information, a point of departure for other, new information. Indefinites are by definition not established. Furthermore, transitive verbs have agents, and agents are usually preferred over patients as topics.

Thus in both Cayuga and Selayarese, all arguments that commonly function as topics are obligatorily represented by morphologically bound pronouns in every clause.

3.2. Discourse-level topicality: Point of view

Although subject selection in accusative languages is normally correlated with such features as person and agency, the ultimate choice is not simply the result of a mechanical calculation of these features in each clause. On occasion, patients of transitive events can function as subjects in languages
like English. The deviation from usual subject choice is motivated by the structure of larger segments of discourse.

Subjects establish an orientation. Speakers do not ordinarily shift this orientation with each new clause; they tend to establish a topic or point of view and retain it over a certain stretch of discourse. (See among others Givón, 1983.) This topic may not function as an agent in every single clause of the general discussion. It may even function as the patient of some event involving another agent. At this point the speaker must decide whether to shift the general orientation of the discourse to the vantage point of the other agent, or to continue from the vantage point of the original topic. Constructions like English passives permit speakers to maintain topic continuity by categorizing patients as subjects.

Because speakers normally retain the same topic over a series of clauses, hearers expect topic continuity until there is a reason for a reorientation. When there is a shift in vantage point or scene, this is usually indicated not only by a shift in subject, but by other devices as well that overtly introduce new topics and alert hearers to the changes.

Cayuga and Selayarese both contain ample stylistic devices for alerting hearers to discontinuity and for foregrounding and backgrounding primary participants. They contain special presentative constructions to introduce new participants. They contain devices such as shifting constituent order, demonstratives, and derivational morphology to indicate the discourse roles of participants. In both languages, full noun phrases are used to alert hearers to topic shifts.

Consider for example the Cayuga passage in (47) from the cosmology legend. A woman has fallen through a hole in the sky, and as she falls, numerous thoughts go through her mind. Suddenly she sees some birds flying around. The discourse topic then shifts from the woman to the birds. The shift is indicated by several devices. The initial particles in the first line signal a reframing. The birds, who had just been introduced (‘and then she noticed some birds’) reappear at the beginning of the second line. The ordering of this noun for ‘birds’ early in the clause, before the verb, serves to foreground it, indicating that it represents significant new information, in this case, a shift in topic. The proximate demonstrative *ne'kye* ‘this’ indicates that the birds occupy center stage.

(47) Cayuga

*Ne'ne'one ne'kye*

it is the now this
t?ite?sho?-? teyotiya?towéhtph atkatiya?to-wéht
birds they are thinking they thought

ne? no ne? ne-kye akenat?enyeté?
the now the this they should try

aposakotiy?atakó?
they would help her

'Now then at this time these birds were thinking that they should try to help her.'

On some occasions, events may imply the involvement of agents that are peripheral to the discourse as a whole. Cayuga contains devices for conveying such events by means of verbs whose only argument is the patient. The passage cited above continues:

ahkwikwa? ne-kye a-yakó-nqunya?-k
that not this OPT-FEM.PAT-hurt-PUNC
that not this she would get hurt

ne? k?ishé ne-kye ne? a-yakó-nyo-?.
it is maybe this that OPT-FEM.PAT-kill-PUNC
or else this that she would get killed
'so that she would not get hurt or get killed.'

The verbs 'hurt' and 'kill' normally involve agents, but in this context their identity was unknown and not worthy of mention.

Cayuga thus contains not only obligatory mention of all clause topics in every verb, it also contains elaborate devices for foregrounding and backgrounding participants according to their roles within the discourse. Sets of particles identify major discourse breaks, shifts in scene. Full noun phrases and word order identify important new participants. Demonstratives point out the locus of center stage. The lexicon and verbal morphology permit the elimination of peripheral agents from consideration as topics.

Selayarese contains similar devices for tracing discourse topics. Significant new participants are commonly introduced with a presentative construction based on rie? 'there is/was'. A story about a policeman opened as follows.

(48) Selayarese presentatives
Ri táuy 1979, rie? pulisi ap-pa-katianay
in year 1979 there is policeman INTR-CAUS-pregnant
The story continued as the policeman promised to marry the pregnant woman, then went home, ostensibly to get money for the wedding. Once home, however, he married someone else. The pregnant woman’s relatives were outraged. When three of them came forward to seek revenge, they were introduced as follows.

(49) Selayarese presentatives

Sängên-na rieʔ-mo tálлу,
finally-3.POSS there·is-PERF three

totarína tàngonŋ laʔ-bálasa.
teenager guarantee FUT-INTR-avenge

‘At last there were three of them, teenagers who guaranteed they would take revenge.’

The teenagers became the new topic of the narrative. They awaited the policeman outside of a movie theater and attacked him.

Shifts in scene and topic are often signaled in Selayarese, as in other languages, by the use of full noun phrases. Once established, discourse topics are referred to by pronouns only, until there is a shift to another topic, at which point a full noun phrase is again used. Note the appearance of the noun phrases underlined in the passage below. The speaker had just said that after about a month at home the policeman returned to Selayar.

(50) Selayarese full noun phrases for topic shift

La-lâygeréʔ-na to-tóá-na to-bahinéŋjo,
3.ERG-hear-3.POSS who-old-3.POSS who-woman-the

lako rieʔ-mo, muliam pulisiŋjo,
that there·is-PERF return policeman-the

mâŋe-mo-i póle la-kutâʔnaŋ,
go-PERF-3.ABS again 3.ERG-ask

sikuuraía naʔráʔraʔ-i ?-bóntiŋ.
when when-want-3.ABS INTR-marriage

Mîŋka pulisiŋjo to-rieʔ-mo-i bahinén-na
but policeman-the who·there·is-PERF-3.ABS woman-3.POSS
ka ri maŋe-na ri kampón-na,  
because at go-3.POSS to village-3.POSS  

gélé-i maŋe ŋéra doe? pa?bóntiŋ  
not-3.ABS go ask-for money wedding  

mĩŋka maŋe-i ?-bóntiŋ.  
but go-3.ABSINTR-marry  

‘When the woman’s parents heard that the policeman had returned,  
they again went to ask him when he wanted to get married. But  
the policeman already had a wife, because when he went to his  
village, he didn’t go to ask for wedding money, but he went to get  
mARRIED.’  

In English, patients of transitives are often encoded as subjects when the  
agent is indefinite or nonhuman. English speakers thus rarely say A car hit  
me or A dog bit me. Speakers are more likely to assume their own point of  
view as a point of departure: I was hit by a car or I got bitten by a dog.  
Selayarese has no equivalent construction. In a normal context, ordinary  
transitive verbs would be used.  

(51) Selayarese transitives with nonhuman agents  

La-lúpuru-a óto.  
3.ERG-hit-1.ABS car  
it hit me car  
‘I was hit by a car’ = ‘A car hit me’  

La-kókkoʔ-a ?ásu.  
3.ERG-bite-1.ABS dog  
it bit me dog  
‘I was bitten by a dog’ = ‘A dog bit me’  

Since person is clear from the pronouns, no information is lost.  
Noun phrases can be foregrounded if they are of unusual importance.  
The sentences below with initial noun phrases would be used contrastively,  
as in ‘It was a car that hit me, not a truck’ or ‘It was a dog that bit me, not  
a bear’, or as the answers to questions ‘What hit you?’ or ‘What bit you?’.
Like Cayuga, Selayarese contains a device for eliminating unimportant agents from the set of primary participants. A derivational prefix ta- forms verbs indicating lack of control. Some non-control verbs have control counterparts.

(52) Selayarese fronting for contrast

\[ ðto \ \text{la-lúpuru-}a \]
\[ \text{car} \ 3.\text{ERG-hit-1.ABS} \]
\[ \text{car it hit me} \]
\[ 'A \text{ car hit me.' (not some other vehicle) } \]

\[ ãsu \ \text{la-kókko-}a \]
\[ \text{dog} \ 3.\text{ERG-bite-1.ABS} \]
\[ \text{dog it bit me} \]
\[ 'A \text{ dog bit me.' } \]

The prefix is derivational. It forms new lexical items. Many verbs containing the prefix have no counterparts without it.

(53) Selayarese non-control verbs with control counterparts

\[ \text{péla?} \quad \text{‘throw (something) away’} \]
\[ \text{tappéla?} \quad \text{‘get lost’} \]
\[ \text{pabánka} \quad \text{‘surprise (someone)} \]
\[ \text{ta?bánka} \quad \text{‘be surprised’} \]
\[ \text{sámbar} \quad \text{‘catch (someone or something)’ (as string catching foot)} \]
\[ \text{tassámbar} \quad \text{‘trip’ (intransitive)} \]

The non-control verbs above are intransitive. (The glottal stop, which assimilates to following voiceless consonants, is an intransitive marker.) The absolutive pronouns that appear with these derived intransitives represent semantic patients.

(54) Selayarese non-control verbs without control counterparts

\[ \text{ta?béssolo} \quad \text{‘slip’} \quad \text{(no *béssolo)} \]
\[ \text{ta?dó?do} \quad \text{‘be sleepy’} \quad \text{(no *dó?do?)} \]
\[ \text{tattó?ro} \quad \text{‘stumble’} \quad \text{(no *tó?ro)} \]

The non-control verbs above are intransitive. (The glottal stop, which assimilates to following voiceless consonants, is an intransitive marker.) The absolutive pronouns that appear with these derived intransitives represent semantic patients.

(55) Selayarese absolutive patients of non-control verbs

\[ \text{ta-p-péla-a} \quad \text{ta-?-béssolo-s} \]
\[ \text{NON.CONTROL-INTR-lose-1.ABS} \quad \text{NON.CONTROL-INTR-slip-1.ABS} \]
\[ 'I \text{ got lost}’ \quad \text{‘I slipped’} \]
The prefix *ta-* can function as a discourse device for eliminating unimportant agents by creating lexical items that do not include these agents among their core arguments. The effect can be similar to that of passivization in English.

(56) Selayarese counterparts of English passive

\[
\begin{align*}
\text{Selayarese} & \quad \text{English} \\
\text{ku-pállu-i} & \quad \text{ta-p-pállu-i} \\
1.\text{ERG-(TR)cook-3.ABS} & \quad \text{NON.CONTROL-INTR-cook-3.ABS} \\
\text{I cooked it} & \quad \text{It is cooked} \\
\text{ku-sássa-mo-i} & \quad \text{ta-tó?ro-a} \\
1.\text{ERG-(TR)wash-PERF-3.ABS} & \quad \text{NON.CONTROL-INTR-stumble-1.ABS} \\
\text{I have washed it} & \quad \text{I stumbled} \\
\end{align*}
\]

The effect of this device can be seen by comparing the two sentences below containing relative clauses. In the first, the agent is overt. In the second, the agent is bypassed by the use of the derived verb ‘get stolen’.

(57) Selayarese use of non-control for backgrounding agents

\[
\begin{align*}
\text{Selayarese} & \quad \text{English} \\
\text{ku-úppa-mo-i} & \quad \text{ásu (nu)-la-lúkka?-ínjø i-Báso.} \\
1.\text{ERG-find-PERF-3.ABS} & \quad \text{dog that-3.ERG-steal-the Mr.-Báso} \\
\text{I have found it} & \quad \text{dog that he stole the Mr. Báso} \\
\text{ta-s-sássa-mo-i} & \quad \text{ta-tó?ro-a} \\
\text{NON.CONTROL-INTR-wash-PERF-3.ABS} & \quad \text{NON.CONTROL-INTR-stumble-1.ABS} \\
\text{It has been washed} & \quad \text{I stumbled} \\
\end{align*}
\]
kuʔúmpa-mo-i ásu (nu)-taʔ-lukkaʔ-iŋjo.
1.ERG-find-PERF-3.ABS dog that-PAT-INTR-steal-the
I have found it dog the that was stolen
'I have found the dog that was stolen.'

4. CONCLUSION

The motivation behind the codification in so many languages of the discourse role of clause topic into a grammatical category of subject is clear. Topic-comment structures are among the most frequent discourse patterns in any language, and identification of the topic, the starting point of a predication, is crucial to its understanding.

Yet neither Cayuga nor Selayarese has grammaticized a subject category in any area of the language. Their formal case marking systems reflect the grammaticization of subtly different distinctions from the clause topicality grammaticized in accusative systems, and their syntactic patterns are sensitive to the distinctions of these systems or none at all. Cayuga grammatical constructions are based either on agent and patient categories or exhibit no case restrictions whatsoever. Selayarese constructions are based on ergative and absolutive categories, control, or no case category at all. The lack of a grammaticized subject category in these languages is not an accident. All participants that function as clause topics are already obligatory represented in every clause by overt morphologically bound pronouns. The features that usually enter into the selection of subjects in accusative languages, namely person and agentivity, are systematically distinguished on the pronouns.

In situations where the clause topic does not correlate with usual person, agentivity, and animacy preferences, other factors guarantee that topicality remains clear. Morphological devices are used to eliminate nontopical agents from the set of core arguments of a predicate. Cues to discourse-level topicality abound in these languages in the systematic use of demonstratives and word order, as well as in the distribution of full noun phrases. As in all languages, hearers generally assume that the topic will remain constant until it is overtly shifted. Although topics themselves are not expressed by a single grammatical category in these languages, shifts in topic, situations worthy of special attention, are signalled by a rich set of devices.

Obviously topicality does not play a lesser discourse role in Cayuga and Selayarese than in languages like English. Cayuga and Selayarese speakers
normally speak as coherently as speakers of other languages, exhibiting the same kinds of topic continuity. The patterns have simply not been formally grammaticized to the same extent. Full grammaticization of a special subject category would be superfluous, given the shapes of the existing grammars. There is thus considerably less motivation for its development. A full understanding of the processes of grammaticization must involve not only a knowledge of the possible sources of grammatical markers and categories, but also of the factors that motivate their development, especially their potential interaction with the grammatical systems in which they emerge. Such interaction is bound to be complex, involving multiple factors and varying degrees of motivation or hindrance. Yet as we discover more about these interactions, we have much to gain in our overall understanding of the forces that mold grammatical systems, of why languages are as they are.

ABBREVIATIONS

AGT     agent
CISLOC cislocative
DIM     diminutive
INCH    inchoative
INTER   inferential
N       neuter
OPT     optative
PAT     patient
SEMI-REFL semi-reflexive

REFERENCES


