Proceedings of AFLA 7

The Seventh Meeting of the Austronesian Formal Linguistics Association

Edited by Marian Klamer

Vrije Universiteit Amsterdam
Department of Linguistics
2000
Proceedings of AFLA 7

The Seventh Meeting of the Austronesian Formal Linguistics Association

Held at the Vrije Universiteit Amsterdam
May 11-13, 2000-07-06

Edited by
Marian Klamer

Vrije Universiteit Amsterdam
Department of Linguistics
2000
ACKNOWLEDGEMENTS

We gratefully acknowledge the funding received for the conference and the proceedings from:
The Dutch Research Foundation (NWO)
The Royal Dutch Academy of Arts and Sciences (KNAW)
The International Institute of Asian Studies, Irian Jaya Studies (ISIR), Leiden
The Centre of Non-Western Studies (CNWS) of Leiden University
The Holland Institute of Linguistics
The Faculty of Arts of the Vrije Universiteit Amsterdam

The following persons helped to organise the conference:
Arke Jongkind, Wilco van den Heuvel, Nanette Huijs, Rob Goedemans

The following persons helped to set the program:
Felix Ameka, Max Planck Institute Nijmegen/Leiden University, Gert Booij, Free University Amsterdam,
Lisa Cheng, Leiden University, Crt Cremer, Leiden University,
Mirjam Emestus, Free University Amsterdam, Rob Goedemans, Leiden & Utrecht University,
Gertjan Postma, Leiden University, Johan Rooryck, Leiden University, Hein Steinbauer, Leiden & Nijmegen
University, Ruben Stoel, Leiden University, Rint Sybesma, Leiden University, Arie Verhagen, Leiden
University, Lourens de Vries, Free University Amsterdam, David Wilkins, Max Planck Institute Nijmegen

© The Authors

TO ORDER:
Send a cheque made out to
Vrije Universiteit, Faculteit der Letteren to:
AFLA VII Proceedings
c/o Vrije Universiteit
Secretariaat Faculteit der Letteren
Kamer 10A14
De Boelelaan 1105
1081 HV Amsterdam
The Netherlands

Price per volume: US$ 25 (postage included)
We cannot accept credit cards
Preface

This volume consists of papers presented at the seventh meeting of AFLA
(Austronesian Formal Linguistics Association), held at the Vrije Universiteit on May

For the first time in the history of AFLA, this meeting was held outside the
North-American continent, and contained contributions by speakers from eleven
different countries: New Zealand, Australia, Indonesia, Brunei Darussalam, Taiwan, the
USA including Hawaii, Canada, the UK, France, Germany, and The Netherlands.

Apart from the languages that are traditionally well-represented at Austronesian
conferences, we were happy to see that the program also contained work on relatively
small or lesser described languages, such as the minority languages of Taiwan, North-
West Borneo, Eastern Indonesia, Papua and Oceania.

Special themes of this conference were Iconicity and Argument marking. The
papers in this volume show that the program covered a broad range of subdisciplines --
from discourse grammar, phonology, morphology, syntax, to semantics -- and that the
authors are working within various theoretical frameworks. But despite the obvious
differences in expertise, interest and background, the atmosphere on the conference was
typically AFLA: lively and constructive, with an average rate of attendance of about
80%. The papers in this volume deserve the same rate of attention.

This meeting has again furthered the unwritten mandate of AFLA to encourage
the formal study of Austronesian languages, especially work by speaker linguists and
junior scholars. Six scholars presented analyses of their native language, and more than
half of the 45 participants subscribed as ‘student’. This suggests that the future of
Austronesian linguistics looks very bright indeed.

The eighth edition of AFLA will be held in the spring of 2001 at the Massachusetts
Institute of Technology (MIT) in Boston, USA. The principal organiser will be Ileana
Paul.

Marian Klamer, Vrije Universiteit Amsterdam

Proceedings of previous AFLA meetings:

A Selection of the papers of AFLA 2, in 1995 is published as:

The proceedings of AFLA 3 and AFLA 4 in 1996/1997 are published as:
Pearson, Mathew (ed.). 1998. Recent papers in Austronesian Linguistics. UCLA

The proceedings of AFLA 5 in 1999 are published as:
Toronto Working Papers in Linguistics.
# Table of Contents

Gabriele Heike Cablitz  
Nominalisation of verbal clauses in Marquesan (Oceanic, French Polynesia) ........................................... 1

Adrian Clynes  
Phonological structures and expressiveness:  
The role of iconicity in ‘the emergence of the marked’ .......................................................... 15

William D. Davies  
Against long movement in Mauroesc .......................................................... 33

Alexandre François  
Vowel shifting and cloning in Motlau: Historical explanation vs. formal description........ 49

Madocyn Kishock  
Transitivity alternations in Rotumian ........................................................................ 69

Thomas B. Klein and Meta Y. Harris  
Fixed segmentism, markedness and faithfulness:  
Nominalising reduplication in Chamorro ........................................................................... 81

Anja Latrouite and Ralf Naumann  
An interpretation of the voice affix /i-/ in Tagalog .................................................... 101

Diane Massam  
Niuean nominalisation .................................................................................. 121

Ulrike Mosel and Jessika Reinig  
Valence changing clitics and incorporated prepositions in Teop ........................................ 133

Simon Musgrave  
Emotion predicates and grammatical functions in Indonesian ........................................ 141

Ileana Paul  
Clefts vs. pseudo-clefts in Austronesian .............................................................................. 155

Phil Quick  
A non-linear analysis of vowel harmony and vowel harmony blocking in Pendau ............................................ 173

Charles Randriamasimanana  
Malagasy. binary branching and null subjects .................................................................... 193

Der-Hwa V. Rau  
Word order variation and topic continuity in Atayal ...................................................... 211
Ger P. Reesink
Austronesian features in a linguistic area .................................................. 231

Li-May Sung
Nominalization in Rukai and Amis ............................................................... 245

Adam Ussishkin
Fixed prosodic effects in Austronesian: An Optimality-Theoretic account .... 259

William A. Foley
Categorial Change in Oceanic Languages:
First Contact on the North New Guinea Coast ........................................... 271
Word Order Variation and Topic Continuity in Atayal

Der-Hwa V Rau

Providence University
E-mail: dhrau@pu.edu.tw

1. Introduction

Previous studies on word order variation have indicated that a correlation exists between the frequency of each word order type in a given language and the factors predicting the choice of one order or another on a particular occasion in that language. This correlation can be established by quantitative approaches to discourse analysis. However, the majority of the text counts have not considered possible skewing of the data, failed to distinguish the truly important factors from the minor ones, and did not consider the possibility of interactions among factors. In this study, a multivariate analysis is used to address these problems.

This study attempts to investigate how discourse functions account for the VS/SV word order variation in Atayal, an Austronesian language spoken in Taiwan, based on a variation approach to discourse analysis (Myhill 1992, Schiffrin 1994). GoldVarb 2.0 (Rand & Sankoff 1990) is used to analyze all the factors accounting for the word order alternation.

A pilot study (Rau 1997b) of word order variation in Atayal indicated topicworthiness was the only factor that could account for word order variation. However, that analysis was based only on one narrative text with 204 tokens. Although one cannot put too much weight on the results of the pilot study due to the idiosyncratic nature of a single text, the methodology points to a promising direction for future study on word order variation. Thus the current study utilizes four times as much data collected from the Mstbon community, where Squiq Atayal is spoken. The results are parallel to those obtained for Ute (Givón 1983) and Chamorro (Coorman 1983) using a different method.

The organization of the paper is as follows. The brief introduction is followed by a literature review of previous discourse studies on word order variation and studies on the grammar and discourse of the Atayalic languages. After the methodology of the study is laid out and the analytical categories are defined and illustrated with examples from the texts, the results and a discussion of the factors accounting for word order alternation between VS and SV are presented.

---

1 For detailed information on the social stratification of the community, phonological variation, and sound change, please refer to Rau (2000).
2. Literature review

2.1. Word order variation

Previous discourse studies on referring terms have shown sequential dependencies across clauses (Ariel 1990, Fox 1987, Givón 1989, Tomlin 1987). The first mentions of a referent tend to be indefinite and explicit, while the second mentions tend to be definite and less explicit. The main functional factor underlying this distribution is information status and accessibility (Ariel 1990. Chafe 1987. Prince 1981).

Textual evidence from Chamorro (Cooreman 1983, 1992), Tagalog (Fox 1985), and Ute (Givón 1983), among others, suggest that word order variation can be explained in terms of referential continuity and thematic continuity. The unmarked VS order correlates highly with referential continuity and is more common in paragraph medial and final position, whereas the marked SV order is used to mark topic discontinuity and is more commonly found when the thematic unity of the paragraph is disrupted.

Myhill (1992:164-215) reviews the results of a number of studies of word order variation in individual languages, focusing on the order of subject, object and verb, and proposes the following two generalizations on sequencing. First, “temporal sequencing”, or an equivalent for Hopper’s (1979) “foregrounding”, correlates strongly with VS order in strongly VS languages. This finding echoes Hopper’s (1987:471) statement, “Verb-initial clauses narrate, noun-initial clauses describe.” Second, a general discourse principle in languages with a high proportion of VS order can be summarized with the following statement: “Putting the most important new information at the beginning of the clause.”

His findings are consistent with Herring’s (1990) universal pragmatic word order pattern, which claims that marked or contrastive topics (or shifted topics) tend to go in initial position in all languages. Other studies of basically verb-initial languages (Payne 1995, Longacre 1995) also suggest that preverbal position is associated with arguments which are more or less “unexpected”.

Cumming (1997) offers an explanation for the association between the discourse functions of predicate-initial order and patient-prominent syntax in Western Austronesian languages. Specifically, parameters related to the semantics of events (including perfective aspect and telicity) and parameters related to the semantics and pragmatics of noun phrases (including definiteness and referentiality of the patient) are shown to correlate in several languages, corresponding to Hopper & Thompson’s (1980) construct of “high discourse transitivity”. In order to ensure the comparability of future studies, Cumming proposes two checklists listing factors of discourse functions and morphosyntactic features relevant to the analysis of constituent order alternations.

Quakenbush (1992) draws attention to the variation of word order by discourse type. He claims that the incontrovertible verb-initial status of Austronesian languages is
established by narrative-based discourse analysis. However, different text types were found to present differing "basic orders", defined as "textual frequency" according to Hawkins' criteria (1983:13). The basic verb-initial word order in Agutayen narrative texts presents a mirror image of that in expository texts. In other words, the "subject-like nominal" (these ang NPs) precede the verb more often in expository texts. Longacre (1982:484) attributes this alternation to the topic-comment structure of expository discourse. In reaction to the narrative bias tradition of discourse analysis, Cumming (1997) also calls for future research to take up questions of constituency and argument structure in conversational discourse.

In summary, in strongly VS languages, the unmarked VS order is associated with topic continuity, whereas the marked SV order is associated with topic discontinuity. Two types of quantitative methodology have been proposed by Givón (1983) and Myhill (1992). The former measures degrees of topicality by counting the referential distance, persistence (decay), and potential interference/ambiguity of the third person referents in narratives. The latter, on the other hand, applies statistical tests to distinguish the truly important factors from the minor ones and considers the possibility of interactions among factors in word order variation.

2.2. Word order in the Atayalic languages

Previous studies on the grammatical structure of the Atayalic languages agree on the primarily verb-initial status of the language group but differ in terms of the definitions of S and O and "basic" word order. In the following review of these studies, the discussion begins by summarizing the generalizations on word order of different dialects and ends with a comparison of the two major views.

2.2.1. Squilq Atayal

According to Egerod's (1966) description, nouns usually follow verbs. However, if the sentence contains a preverb (including nonconjugatable quasi-verbs, verb particles, modal adverbs, and auxiliaries) with or without a main verb, the pronouns immediately follow the preverb. He did not take an explicit position on the issue of basic word order.

Chen & Lin (1985) propose that the general word order in Atayal is subject follows verb and object follows subject. Here the S and O refer to semantic subject and object.

Rau (1992) describes the word order in terms of the order of subject and predicate, an equivalent of Halliday's (1985) theme and rheme, or the topic-comment structure. The predicate normally comes before the subject in the unmarked cases. However, when the subject precedes the predicate, special attention is drawn to the subject. The subject may be followed by the particle ga or a pause to form a topic separate from the rest of the...
sentence. In other words, the often-used commentary construction in Atayal, as suggested by Legerod (1988), exhibits the form of 'as for A, (there is) B', where A is the theme, B the rheme. In a discourse analysis of Atayal narrative, Rau & Grimes (1994) claim that the basic order in Atayal is VSO, where S and O represent the typical semantic agent and patient of a transitive verb respectively. We suggest word order variation in Atayal may correlate with the referentiality and definiteness of the Os. In other words, non-referential indefinite Os generally co-occur with VOS order while referential definite Os co-occur with VSO order. Our findings support Hopper and Thompson's "high discourse transitivity" as mentioned previously.

Huang (1993) claimed VOS is the basic order of Wulai Atayal. The S refers to a semantic agent rather than a grammatical subject. She noticed a correlation between Atayal word order and its voice system. VOS order was found to generally co-occur with verbs marked by m-/m- affixes whereas VSO order with verbs affixed with -un, -an, or s-. However, in a later frequency count of m-clauses and non-m clauses, Huang (1994) observed that non-m clauses had a higher frequency in texts than m-clauses and claimed that non-m clauses are canonical transitive clauses both syntactically and semantically. This seems to indicate she has changed her view of the canonical word order from VOS to VSO since the latter is associated with the more frequent non-m clauses.

2.2.2. C'uli' Atayal

Li's (1995, 1996) studies of Mayrinax and Skikun syntax claim that verbs (or predicates) generally occur in sentence initial position and are followed by the subject and the object. His S and O refer to semantic subject and object.

In their studies of Mayrinax syntax, Mei (1994), Chang (1995), and Huang (1995) all point to the same canonical word order in Mayrinax Atayal, i.e., VOS. The S and O are no longer semantic agent and patient respectively, but rather the grammatical subject and object.

2.2.3. Seediq

Both S. Huang and Su investigated the Nakahara dialect of Seediq (i.e., Tkinda). Huang's (1997) findings can be summarized into the following five points. First, the basic word order in Seediq is VOA. The definition of the valency roles follows Dixon (1979). A and O refer to the typical agent and patient of a transitive verb and S the single argument of an intransitive verb. Second, the preferred argument structure (PAS) in Seediq discourse appears to display what Du Bois (1987) termed one Lexical Argument Constraint and the Non-Lexical A Constraint. Third, Seediq displays ergative patterning in both grammatical and pragmatic dimensions of PAS. However, S=A links outnumber
S-O links in terms of topic continuity. Fourth, word order and focus in Seediq are interdependent and mutually predictive. The unmarked word order for Agent Focus (AF) is VOA, while that for Non-Agent Focus (NAF) is VAO. Finally, Seediq was not found to be what Cooreman, Fox and Givón (1984) termed “discourse ergative” since the measure of topic continuity did not exhibit any difference between AF and NAF.

Su (1997a) observes the word order in Seediq is VOS, or VSO when bound pronouns occur. The participant in focus is generally found in sentence-final position and the semantic roles of the participants are mostly determined by word order. Here the S refers to a grammatical subject. She (1997b) further argues that the pivots in Seediq display a grammaticization process changing the topic nominals into the subject-like ones.

Chang (1997) proposes that Seediq is a verb-initial (or predicate-initial) and subject-final language. But pronominal subjects must attach to the sentence-initial verbal elements. Again, his use of subject refers to a grammatical subject.

2.2.4. VSO, VOS or SVO?

The difference in the two proposed basic word orders, i.e., VOS vs. VSO in the Atayalic languages, on the surface seems to lie in how S and O are defined. However, when examined closely, the two views actually point to the same conclusion. If S and O refer to semantic roles of typical agent and patient in a transitive clause, the basic word order is VOS (or VOA) or VSO (or VAO). The former is the unmarked order for Agent Focus while the latter for Non-Agent Focus. Notwithstanding the word order variation, the focused element occurs in the sentence final position. If, on the other hand, S and O are defined as grammatical subject and object respectively, the canonical word order is VOS, where S is the pivot, the grammatical subject, or the focused element. In other words, the two interpretations do not really differ in predicting where the verb and the focused element occur. They both agree on the verb or predicate initial status of the language group and the sentence final position of the focused element.

Although the basic word order of Atayal has been well established, very little investigation has been conducted on the word order alternation between VS and SV, except for Rau's preliminary study (1997). In fact, the preposed topics/subjects are a common feature in not only the Atayalic languages but in Formosan languages in general (Starosta 1988). Thus the current study aims to fill in the gap by investigating the discourse factors that have a significant effect on word order variation in Atayal. S and V here refer to the grammatical subject NPs (or the focused argument) and predicate respectively.
3. Methodology

3.1 Research questions and hypotheses

The hypothesis tested in this paper is that word order variation (pre-verbal vs. post-verbal subject) in Atayal can be accounted for by the following factors: the information status and topic continuity of the subject, the grammatical roles (i.e., agent, patient, beneficiary, etc.) of the subject and voice alternations, nominalization of the subject derived from dependent clauses, and the gender and age of the speakers.

First, the word order variation between pre-verbal and post-verbal subject in Atayal is hypothesized to correlate with the information status and topic continuity of the subject. In terms of the information status of the subject, we would predict post-verbal subject to be expected while the preverbal subject unexpected, based on the findings of Cooreman (1992) and Hopper (1987), as discussed in 2.1. Furthermore, following Givón’s hierarchy, as presented as follows in (1), the syntactic devices at the top involve topic continuity and less surprise, while those closer to the bottom involve less topic continuity and higher surprise.

(1) zero anaphora > unstressed /bound pronouns or grammatical agreement > stressed/independent pronouns > right dislocated NP’s > simple definite NP’s > left dislocated NP’s > indefinite NP’s > Y-movement > cleft/focus constructions

In other words, VS order is associated with previous mentions, paragraph medial and final, definiteness, referentiality, inexplicitness, more topicworthiness, foregrounding, and topic continuity, while the SV order is associated with first mentions, paragraph initial, indefiniteness, non-referentiality, explicitness, less topicworthiness, backgrounding, and topic discontinuity.

Second, the grammatical roles (i.e., agent, patient, beneficiary, etc.) of the subject and voice alternations (active, ergative, etc.) are correlated with word order variation. Previous studies indicate co-occurrences between VOS order and AF or active and VSO order with NAI or ergative, but no study has investigated the relationship between voice alternations and the pre-verbal S.

Third, nominalization of S derived from dependent clauses is tested for its relationship with word order variation because nominalization is sensitive to the demands of discourse, as pointed out by Hopper and Thompson (1980, 1984). Nominalization might be associated with a background rather than a foreground.

Finally, the sex and age factors are tested for preferences of word order.

3.2. Data

The data consist of 10 narrative texts of the Squilq dialect of Atayal, spoken in
Mstbon (Rui-yan in Chinese) community, in Ren’ai Village, Nantou County, Central Taiwan. The texts were tape-recorded and transcribed by Pastor Batu Temu, a native speaker of Atayal in his 30s, who was trained to do the interviews and transcribe the data. He was raised in the community and has established good relationships with the local residents. The distribution of the demographic background of the speakers is presented as follows in Table 1.

Table 1. Demographic background of the 10 speakers

<table>
<thead>
<tr>
<th>Name</th>
<th>Sex</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lawa Naway</td>
<td>F</td>
<td>32</td>
</tr>
<tr>
<td>Sayung Nawa</td>
<td>F</td>
<td>74</td>
</tr>
<tr>
<td>Temu Nakaw</td>
<td>M</td>
<td>73</td>
</tr>
<tr>
<td>Temu Syat</td>
<td>M</td>
<td>67</td>
</tr>
<tr>
<td>Temu Piling</td>
<td>M</td>
<td>66</td>
</tr>
<tr>
<td>Kumu Tusang</td>
<td>F</td>
<td>58</td>
</tr>
<tr>
<td>Walis Nabu</td>
<td>M</td>
<td>54</td>
</tr>
<tr>
<td>Abey Toci</td>
<td>F</td>
<td>45</td>
</tr>
<tr>
<td>Ciwas Nabu</td>
<td>F</td>
<td>39</td>
</tr>
<tr>
<td>Terow Nasan</td>
<td>M</td>
<td>35</td>
</tr>
</tbody>
</table>

The texts were divided into 22 paragraphs and 840 clauses. A paragraph boundary is determined by a change of plot, topic, location, or time. Following S. Huang (1997), the clause is defined as consisting of a verb (or predicate nominal or adjective) and its core argument NPs, but excluding oblique nominals.

All the preverbs (tense/aspect/mode markers, negators, adverbs) are coded as V, following Starosta’s (1988:555) AMV (auxiliaries as main verbs) analysis and Jeng’s (1997:213-281) classification of Bunun verbs.

Both the occurrence of S and zero anaphoras were counted. The position of zero anaphoras were coded as following the verb because all the bound pronouns occur after the verb.

Following Young & Bayley’s (1996:261) coding procedure for VARBRUL, the dependent variable is coded in factor group 1, then the co occurring contextual features are coded in factor groups 2-15. Each factor within a factor group is assigned a unique one-character identifier in parentheses. Each factor group should be orthogonal or independent from all others and all the factors in each factor group are exhaustive (Guy 1993).

3.3. Coding

The following is a list of the analytical categories examined in this study. The
coding sheet is first presented, followed by definitions and examples from the texts to illustrate the coded categories.

3.3.1. Coding sheet
(1) Word order variation: VS (v), SV (s)
(2) First mentions: first-mentions (f), recent second mentions (s), resumptive (r), text-structure boundary (b).
(3) Position within paragraph: initial (i), medial (m), final (f).
(4) Definiteness: definite (d), indefinite (i).
(5) Referentiality: referential (r), non-referential (n).
(6) Explicitness: simple NP (i), modified NP (e), zero (z).
(7) Topicworthiness: first person pronoun (f), second person pronoun (s), third person pronoun (t), proper name (p), human NP (h), non-human animate NP (a), inanimate NP (i).
(8) Temporal sequencing: background (b), foreground (f).
(9) Topic particle in SV clauses: ga (g), none (n), non-applicable (/).
(10) Gender: male (m), female (f).
(11) Age: 30's (a), 40's (b), 50's (c), 60's (d), 70's (e), 80's (f).
(12) Nominalization of S derived from dependent clauses: yes (y), no (n).
(13) Grammatical roles of S: Agent (a), Patient (p), Beneficiary (b), Locative (l), Instrumental (i), Conveyance (c).
(14) Pronouns: free personal pronoun (f), bound personal pronoun (b), demonstrative pronoun (d), interrogative pronoun (i), zero (z), non-applicable (/).
(15) Voice alternations: active (a), antipassive (t), ergative (e).

3.3.2. Analytical categories: definitions and examples
(1) Word order variation
S can occur after the verb, such as “I” and “the fire” in (1a), or before the verb, such as “my father” and “my mother” in (1b).

Abreviations:
1PGEX=first person plural genitive exclusive, 1PGIN=first person plural genitive inclusive, 1PG=first person genitive, 1SD=first person singular dative, 1SG=first person singular genitive exclusive, 1PNEX=first person plural nominative exclusive, 1SN=first person singular nominative, 1SNF=first person singular nominative free, 1SD=first person singular dative, 2SG=second person singular genitive, 3SD=third person singular dative, 3SG=third person singular genitive, 3SNF=third person singular nominative free, AF=agent focus, ASP=aspect, BF=benefactive focus, CF=conveyance focus, CON=conjunction, FUT=future, GEN=genitive marker, IF=instrument focus, IRR=irrealis, LF=locative focus, NEG=negation, NOM=nominative, PAR=particle, PERF=perfective, PF=patient focus, PN=personal name, TOP=topicalization.
Word Order Variation and Topic Continuity in Atayal

(1a) m'abi saku ru ini muyan ktay punaq qasa ru, AF-sleep 1SN CON NEG 1PGEX see-LF fire that CON
    "I fell asleep and did not see the fire."
(1b) yaba maku ga wan mhoqin cipoq saku na. father 1SG TOP PERF AF-die small 1SN yet
    "My father died when I was still young."

(2) First mentions
   S is coded as “first-mention” if it has not been mentioned previously in the discourse,
   “recent second mention” when S has been mentioned in the previous clause in the
discourse, “resumptive” when S is being reintroduced into the discourse after a lapse, or
   “text-structure boundary” when S has been mentioned in the previous discourse, but a
   paragraph boundary intervenes since the prior mention.
   In the following example (2a), first person singular pronoun "I" is illustrated in terms
   of its coding. The free nominative form kuzing is coded as "first mention" at the
   beginning of the narrative in (1). The short bound nominative ku is coded as "recent
   second mention" in (2). After a paragraph boundary, the long bound nominative form
   saku is coded as "text-structure boundary". After a lapse, the long bound nominative
   form saku is coded as "resumptive" in (9) when the S is reintroduced into the discourse.
   Finally, the short bound nominative ku is again coded as "recent second mention". The
   occurrences of short vs. long forms seem to follow the principle of iconicity. The short
   form encodes inexplicitness and recent second mention while the long form explicitness,
   first mentions, and reintroduction into the discourse.

(2a) (1) kuzing niya ga lau maku ga Temu Nakaw.
    1SNF 3SNF TOP name 1SG TOP PN
    "I, my name is Temu Nakaw."
(2) yaba ku na Batu Bokusi.
    Father 1SN GEN PN Pastor
    "I am Pastor Batu's father."
(3) ru kawas maku hiya ga pitu pgan ciwan.
    CON age 1SG 3SNF TOP seven ten three
    "I am 73 years of age."
    [Clauses 4-6, Paragraph boundary]
(7) nwha saku smi rusa nxan,
    PERF.AF-go 1SN AF-put trap life
    "In the old days, I went to set traps."
(8) ru pitu bingi lga, say ta mlaw,
    CON seven day PAR go-LF.IRR.1PG AF-check
    "After seven days, we were going to check (it)."
(9) mha saku.
    Thus 1SN
    "I did that."
(10) mtuleq ku sasan,
    AF-rise 1SN morning
    "I rose in the morning."
    [...] 

(3) Position within paragraph
    S may occur in the paragraph initially, medially, or finally.

(4) Definiteness
    S may be definite or indefinite depending on whether it is identifiable to the hearer.
In Atayal, nouns preceded by qutux "one" or without any determiners are coded
"indefinite".

(5) Referentiality
    S may be referential or non-referential depending on whether it is identifiable to the
speaker.

(6) Explicitness
    S may be coded as a simple NP, as qulih "fish", modified NP, as qulih raran "fish in the
past", qulih misuw qani hiya "fish nowadays", or zero.

(7) Topicworthiness
    The term “topicworthiness” is equivalent to what typologists used to call “animacy”.
Our categories correspond with Silverstein’s Animacy Hierarchy (1976), except that the
kin term is grouped with the human NP instead of proper names. S may occur as first
person pronoun, second person pronoun, third person pronoun, proper name, human NP,
non-human animate NP, or inanimate NP.

---

3 Whaley (1997) suggests that “animacy” is a micrometer because it is only one of the parameters that are
reflected in the so called “Animacy Hierarchy”. The other parameters are sociocentric orientation, empathy,
and definiteness.
4 The result of the pilot study indicated that proper names might be associated with SV order. Therefore,
proper names were separated as a factor to be tested in this study.
(8) Temporal sequencing

S may occur in foregrounded or backgrounded clauses. Foregrounding refers to the thread or backbone of the discourse, which is ordered in temporal sequence and frequently associated with non-agent focus, whereas backgrounding refers to the clauses that merely assist, amplify, or comment on the speaker’s goal and are not ordered with respect to each other, frequently associated with agent focus.

(9) Topic particle in SV clauses

S is marked with a special particle ga for topicalization, reserved for pre-predicate elements, as in (1b) and (2a-1), or none, as yaya nanak “only my mother” without the following topicalizer ga in (9a).

(9a) trang cipok myan ga, uka qa yaba, yaya nanak maki.
    when small 1SG PAR no that father mother only AF-exist
    "From my childhood, I had no father. Only my mother was alive."

In SV order, S may be an L-dislocated NP or a Y-moved NP. In the L-dislocation construction, S precedes a complete clause, with a pronominal element cross-referencing S in the clause, as in (9b). In the Y-movement construction, on the other hand, S precedes a predicate, without a pronominal element cross-referencing S in the predicate, as in (9c).

(9b) maki qutux ryax,
    AF-exist one day
    sami yangu mu,   nwah myan sbyayux tmubux pagay.
    1PNEX younger brother's wife 1SG PERF.AF-go 1PGEX share AF-plant rice
    "One day, my younger brother's wife and I, we went to share the work in rice planting."

(9c) yaya yaya anay uku mpanga.
    that female lett-LF.IRR 1SG AF-carry
    "That female (pig), let me carry."

(10) Gender

The 10 speakers were equally divided between males and females.

(11) Age

The 10 speakers were divided into six age groups: 30's, 40's, 50's, 60's, 70's, 80's.

---

5 This is a homophonic particle ga meaning "when, after", functioning as a conjunction for adverbial clause.
6 The S is treated the same way as A and differently as O in discourse.
(12) Nominalization of S derived from dependent clauses
S may occur in nominalized construction derived from dependent clauses, as in (12a).

(12a) cipok alay qu miyuk uzi qa qoqoq qa,
Small very NOM AF-enter also that spear that-PAR
"The stabbing of the spear was not deep either."

(13) Grammatical roles of S
The grammatical roles of S may be either Agent (13a), Patient (13b), Beneficiary (13c), Locative (13d), Instrumental (13e), or Conveyance (13f).

(13a) ru ana ga ulung su qmalu utux kyan.
CON but PAR fortunate 2SG AF-bestow god heaven
"But fortunately the Heavenly God bestowed grace."

(13b) qingan ku nya ptehuk qalang
support-PF 1SN 3SG FUT-arrive village
"He supported me to get home."

(13c) hazi ku nya sgalu uzi.
maybe 1SN 3SG BF-bestow also
"Maybe He bestowed grace to me."

(13d) biqan ku nya kinlokah.
give-LF 1SN 3SG strength
"He gave me strength."

(13e) ciwan sbiq maku hiyan.
three IF-give 1SG 3SD
"I gave him three (fish)."

(13f) ana ga cyux ku nya snaga kraya.
but PAR ASP 1SN 3SG CF-wait above
"But he was awaiting me above."

(14) Pronouns
Pronouns are coded as free personal pronoun, bound personal pronoun, demonstrative pronoun, interrogative pronoun, or zero. The categorization of pronouns were constructed to test Givón’s scale of topic continuity (Givón, 1983a:17).

The system of personal pronouns in Atayal is presented in Table 2 (cf. Rau 1992, and Huang 1993).


Table 2. Atayal Pronominal System

<table>
<thead>
<tr>
<th>Person/Number</th>
<th>Bound</th>
<th>Free</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nominative</td>
<td>Genitive</td>
</tr>
<tr>
<td>1S</td>
<td>saku-ku</td>
<td>maku-mu-amu-aku-uku</td>
</tr>
<tr>
<td>2S</td>
<td>su</td>
<td>su</td>
</tr>
<tr>
<td>3S</td>
<td>hiya</td>
<td>nya</td>
</tr>
<tr>
<td>1P (inclusive)</td>
<td>ta</td>
<td>ta</td>
</tr>
<tr>
<td>1P (exclusive)</td>
<td>sami</td>
<td>myan-muyan</td>
</tr>
<tr>
<td>2P</td>
<td>simu</td>
<td>manu</td>
</tr>
<tr>
<td>3P</td>
<td>hga-lhga</td>
<td>nha</td>
</tr>
<tr>
<td>1SG+2SN=</td>
<td>misu</td>
<td></td>
</tr>
</tbody>
</table>

The following examples illustrate interrogative pronouns (14a), demonstrative pronouns (14b), and zero (14c).

(14a) son ta nanu la?
go-PF 1PGN what PAR
“What are we going to do?”
(14b) qani ga ciriqan nanu yaba su.
here TOP catch-LP what father 1SG
“Here is what your father caught.”
(14c) nyux mtakuy ru nyux mu sr’un ra.
ASP AF-fall CON ASP 1SG hold down-PF PAR
“After (it) fell, I was holding (it) down.”

(15) Voice alternations: active (a), antipassive (t), ergative (e).

S may occur in active constructions, including one-argument intransitive clause (13a), existential sentence (15a), equational sentence (15b), and adjectival predicates (15c), antipassive constructions (15d), or ergative constructions, including patient focus (13b), benefactive focus (13c), locative focus (13d), instrumental focus (13e), and conveyance focus (13f).

(15a) ru yasa qa nuka yaba mu la
CON therefore that no father 1SG PAR
"Therefore, I have no father."
(15b) lau maku ga Lawa Naway.
Name 1SG TOP PN
"My name is Lawa Naway."
(15c) yaqih balay qaxan myan.
had very life 1PGFX
"Our life was very difficult."
(15d) yaya maku qutux Imanga knan.
Mother ISG one AF-raise 1SD
"My mother raised me as a single parent."

4. Results and discussion
4.1. An initial VARBRUL run
In this section, we simply present the results of the preliminary run. Detailed discussions on the final results are deferred to the next section on recodings and subsequent runs.

Our initial VARBRUL run shows the VS order (750/840, or 89%) is the unmarked word order. Ergative or non-agent focus (36%) occurs more frequently than anti-passive (13%). But the active construction (including one-argument agent focus, existential sentences, equational sentences, and adjectival predicates) is more frequent than either the ergative or the anti-passive construction, constituting 50% of all the voice alternations.

VS order seems to be associated with previous mentions, paragraph medial and final positions, definiteness, referentiality, explicitness, more topicworthiness, foregrounding, and topic continuity, if we compare all the percentages of the factors affecting VS order in each factor group. The factors with the highest percentages associated with VS order are recent second mentions (97%), paragraph medial position (90%), definite (91%), referential (91%), simple NPs (94%), no nominalization of S (89%), first person (96%), foregrounding (93%), and bound personal pronoun (100%).

On the other hand, SV order seems to be associated with first mentions, paragraph initial, indefiniteness, non-referentiality, explicitness, less topicworthiness, backgrounderg, and topic discontinuity, if we compare all the percentages of the factors affecting SV order in each factor group. The factors with the highest percentages associated with SV order are text-structure boundary (42%), first mentions (35%), paragraph initial position (52%), indefinite (26%), non-referential (27%), modified NP (26%), nominalization of S (14%), inanimate NP (21%), backgrounderg (17%), and free personal pronoun (59%).

In terms of social factors, male speakers (91%) seem to have a preference for VS order while female speakers (16%) seem to have a preference for SV order. The oldest group has a preference for SV order, but this number is based only on the speech of one female speaker in her 80's.

However, our quantitative analysis certainly does not stop here because a closer look reveals that only paragraph initial position and independence of personal pronoun indicate
slightly higher percentages of SV order than VS order. Besides, the knock-out factors have to be recoded, and the independence of the factor groups and the information on goodness-of-fit have to be determined before the results can be interpreted. Thus the factors were recoded for subsequent VARBRUL runs.

4.2. Re-coding

Several recodings and subsequent VARBRUL runs have brought us to recode our factor groups as follows:

(1) Word order variation: VS (v), SV (s).

(2) First mentions: first mentions (f), recent second mentions (s), resumptive (r), text-structure boundary (b).

(3) Explicitness: modified NP (m), simple NP and zero (s).

(4) Pronouns: zero and bound personal pronoun (b), free personal pronoun (f), demonstrative pronoun (d), interrogative pronoun (i), non-applicable (l).

4.3. One-level binomial analysis

The results of the one-level binomial analysis are shown in Tables 3.

<table>
<thead>
<tr>
<th>Group</th>
<th>Factor</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) First mentions</td>
<td>First mentions</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>Recent second mentions</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>Resumptive</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>Text-structure boundary</td>
<td>0.11</td>
</tr>
<tr>
<td>(2) Explicitness</td>
<td>Modified NPs</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>Simple NPs and zero</td>
<td>0.61</td>
</tr>
<tr>
<td>(3) Pronouns</td>
<td>Bound personal and zero</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>Free personal</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>Demonstrative</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>Interrogative</td>
<td>0.59</td>
</tr>
</tbody>
</table>

The input probability, that is the likelihood that the post-verbal rule (VS order) will operate irrespective of conditioning factors, is 0.81. First mentions, explicitness and independence of the pronouns are the three factor groups that affect word order variation in Atayal. In terms of first mentions, recent second mentions (0.80) and resumptive (0.73) favor VS order, while first mentions (0.43) and text-structure boundary (0.11) favor SV order. As for explicitness, simple NPs and zero (0.61) favor VS order, whereas modified NPs (0.39) favor SV order. Finally, in terms of the independence of the pronouns, bound
personal pronouns and zero (0.97) and interrogative (0.59) favor VS order, while free personal pronouns (0.06) and demonstrative pronouns (0.26) favor SV order. The results indicate that information status of the S and topic continuity affect word order variation in Atayal.

The three factor groups and all the factors are independent because the error values are mostly below 1.5, the total Chi-square value (5.1041) is within the limit (6.490), and the Chi-square per cell figure (0.2686) is also very low. The result of the Chi-square test indicates goodness-of-fit and the least likelihood that interaction exist among factors.

4.4. Step-up/step-down binomial analysis

A step-up/step-down binomial analysis was conducted to test whether all the three factor groups in the last run contribute to the observed pattern of variation. The probability level is 0.05. The best stepping up and stepping down runs selected all three factor groups. In other words, first mentions, explicitness, and independence of pronouns have significant effect on word order variation in Atayal.

4.5. Discussion

Our hypothesis that word order variation is affected by the information status and topic continuity of the subject is supported. First mentions, explicitness, and independence of pronouns have significant effect on word order variation in Atayal. Recent second mentions, resumptives, simple NPs, bound personal pronouns, zeroes, and interrogative pronouns favor the VS order whereas text-structural boundary, first mentions, modified NPs, free pronouns, and demonstrative pronouns favor the SV order.

The fact that interrogative pronouns occur more frequently in VS order needs explanation. In WH-questions, the interrogative pronouns functioning as subject in action-oriented clauses do not change order, as illustrated in (4-1), (4-2).

(4-1) son ta nanu la?
    go-PF 1PG what PAR
    "We are going to do what?"
(4-2) kutan inu, mama?
    cut-LF where uncle
    "Cut where, uncle?"

The interrogative pronouns occur pre-verbally in cleft/focus construction or state-oriented clauses, such as (4-3).
(4-5) nanu nyux su nqun, Temu?
    what ASP 2SG eat-PF PN
    "What is it that you are eating, Temu?"

Factors that were not found to have any significant effect on the choice of VS or SV include position within paragraph, definiteness, referentiality, topical worthiness, temporal sequencing, grammatical roles, voice alternations, nominalization of the subject, and social factors of gender and age. However, the division of paragraphs, the assignment of definiteness, and the decision on foregrounding and backgrounding are yet to be worked out in a more objective way.

5. Conclusion

The study confirms that word order variation is associated with the information status and topic continuity of the Subject. VS order is associated with old information and topic continuity while SV order is associated with new information and topic discontinuity. In particular, recent second mentions, resumptives, simple NPs, bound personal pronouns, zeroes, and interrogative pronouns favor the VS order whereas text-structural boundary, first mentions, modified NPs, free pronouns, and demonstrative pronouns favor the SV order.

By using a multivariate analysis, GoldVarb 2.0, to distinguish the truly important factors from the minor ones and remove interactions among factors in the design, this study addressed the problems that previous studies of word order variation based on text counts have generally failed to do.

References


