Limitation and variation in the word order of VOS languages

若　山　真　幸

WAKAYAMA Masayuki

1. Introduction

This paper presents a rough sketch of the harmonious/disharmonious head-complement order of VOS languages with a special focus on some Formosan languages spoken in the main island of Taiwan. Basic word order is one of the best components to explore language variation. The orders of S, V, and O can logically yield six word order patterns. In addition, if we focus on the relative order within a phrase - specifically, the relative order V (head) and O (complement) within VP - the variation in word order can be summarized into just two types: head-initial and head-final order. For example, Japanese and English have an SOV and SVO order, respectively and therefore, the former is a head-final, while the latter is a head-initial language. On the other hand, Kayne (1994) proposed that linear order within a phase domain is universally determined by the Linear Correspondence Axiom. In other words, there is no hierarchical difference between Japanese OV and English VO order in syntax. Chomsky, Gallego, and Ott (2017: 233) claim that "if order is only established in the morphophonological component, no syntactic operation can make reference to it." However, in fact, there are strong preferences in word order; SOV and SVO are noticeably most dominant orders among the six orders. As illustrated by Greenberg (1973), there are additional correlations across categories in the head-complement sequences. Furthermore, Cinque (2005) found, contrary to Greenberg's Universal 20[1], that only 14 orders out of the 24 possible orders (combinations of Dem, Num, Adj, and N) within the DP phrase are attested. Why do we enjoy this variation with "semantically vacuous movement" in the sense of Cinque (2018)? Therefore, in this paper, I will focus on verb-initial languages with the VOS word order and examine the correlations of head-complement sequences.

---

[1] When any or all of the items (demonstrative, numeral, and descriptive adjective) precede the noun, they consistently appear in that order. If they follow, the order is either the same or its exact opposite.

2. Geographical and historical backgrounds

The present study focuses on some Formosan languages, spoken in the main island of Taiwan, for two main reasons. Firstly, they are assumed to be the origin of the Austronesian language family, which extends currently from the Southeast Asian islands, such as the Philippines and Indonesia, to the Polynesian islands, covering regions like Papua New Guinea and New Zealand[2]. Secondly, the Formosan languages, estimated to have been separated from the Proto-Austronesian language around 6,000 years ago, retain some unique characteristics of the Proto-Austronesian language such as VOS or VSO word order. Seediq and Amis are Formosan languages[3] spoken by indigenous people of Taiwan. Seediq is categorized as an Atayalic language with a dominant VOS order, while Amis is a genus of East Formosan spoken on the East coast of Taiwan with a dominant VOS or VSO word order. In either case, the verb generally precedes the subject; that is to say, they are head-initial.

3. Basic information of the VOS word order

According to WALS[4], the number of VOS language is only 25, compared with 564 SOV and 488 SVO languages. VOS is found scattered across the world, from the islands of Indonesia to North, Central, and South America areas. Among the six dominant orders, the number of VOS, OVS, and OSV is quite low (e.g. 40 languages in total). This probably stems from the relative order of O and S; the SO order is more common than the OS order.

(1) SOV   564 languages[5]

   SVO   488

   VSO   95

   VOS   **25**

   OVS   11

   OSV   4

   No dominant order 189

WALS also presents patterns for languages with two dominant orders of S, O, and V. Regarding VOS, there are only two alternative orders, as shown below.

---

[2] The geographical range of this language family is considerably broad, including languages such as Hawaiian and the Malagasy language spoken in Madagascar.

[3] Amis is a member of the East Formosan language spoken in the eastern area of Taiwan island while Seediq belongs to the Atayalic language used in the northwest of the island.

[4] The World Atlas of Language Structures (http://wals.info)

[5] http://wals.info/chapter/81

(2) VSO or **VOS**　　　　　　14 languages

　　　SVO or **VOS**　　　　　　8

Among the 14 languages of VSO or VOS orders, seven belong to the Austronesian language family, and there are three languages with the SVO or VOS order pattern. The correlation of VSO and VOS is theoretically predictable because VOS belong to the head-initial group such as SVO and VSO. On the contrary, the alternation of SVO and VOS is not consistent in terms of head-directionality.

　　It should be pointed out that the word order changes in (2) do not occur freely. For example, Seediq exhibits a primary VOS word order, as shown below[6].

(3) a. qnills　　　　　[kalat　　　　　niyi]　　　[ka.emphapuy]

　　　AV[7]-peels　　　pineapple　　　this　　　NOM　　　cook

　　b. qnlis-an　　　　[emphapuy.ka]　　　　　[kalat　　niyi]

　　　peels-GV[8]　　　cook　　　NOM　　　　　pineapple　this

　　　The cook peels this pineapple.'

In (3), the verb *qnills*, which occupies the initial position of the sentence, is followed by the object *kalat nivi*, and the subject *ka emphapuy*. In addition, the demonstrative pronoun *nivi* always follows the noun (*kalat/pineapple*), while the order of the nominative case marker is not consistent. They point out that Truku also allows an SVO order, which is derived from the canonical VOS order. They assume that the SVO order is more costly and complex since it takes more time to respond to the SVO order compared to the VOS order. In (4a) below, the subject is fronted to the initial position of the sentence, with the object following the other two items.

(4) a. [emphapuy o]　　q nills　　　　[kalat　　　niyi]

　　　cook　　TOP　　AV-peels　　pineapple　this

　　b. [kalat　　niyi]　　o　　　　qnlis-an　　　　　[emphapuy]

　　　pineapple this　　TOP　　peels-GV　　　cook

Aldridge (2020) maintains that the alternation between VOS and VSO in Seediq can be

---

[6] This data come from the Truku (Taroko) language, a dialect of Seediq. (cf. Ono et al. (2020)) In this literature, the case is referred to as NOMINATIVE. However, it should be ABSOLUTIVE.

[7] AV stands for Agent voice.

[8] DV stands for Goal voice, which is roughly equivalent to the passive voice with a benefactive meaning in English.

attributed to the definiteness of the object. Specifically, VP fronting in Seediq is possible when the object is indefinite and nonspecific, while it remains in its original position when the object is definite and specific. To support this view, the TOPIC marker is added to the noun in this structure, instead of the nominative case. Furthermore, the subject of SVO receives a strong focus intonation. (cf. Karlsson and Holmer (2011)) In any case, it is important to stress that word order is not arbitrary but is influenced by specific factors or conditions.

3.1 Derived word orders

In theoretical linguistics, it has long been assumed that word order differences are consequences of movement. As mentioned in section 1, SVO is the sole underlying order in the Chomskyan theoretical framework. Therefore, in the case of Japanese, which has an SOV order, there is an additional movement of the object (O) out of the verb phrase (VP) to a higher position. Beyond the VP domain, the constituents of the complement also move to a higher position to establish the desired order. Consider (3b). In Japanese, the complementizer *to* (equivalent to *that* in English) occupies the sentence-final position just because this language is head-final within the traditional grammar. On the other hand, in lines with Kayne (1994), the head-final status of C is only achieved through 'meaningless scrambling' of TP in the sense of Cinque (2018).

(3) a. [Subj [Obj$_i$ [$_{VP}$ V t_i]]]   SOV order
    b. [TP$_i$ [$_{CP}$ [C (=*that*) t_i]]]   TP-complementizer order
    c. [DP$_i$ [$_{PP}$ [P t_i]]]   postpositional phrase

What about the VOS order? In his theory, because Spec is always in the left side of the head, the most straightforward ways it to move VP over S to a higher position.

(4)   VOS: [$_{XP}$ [$_{VP}$ V+O]$_i$ [$_{vP}$ Subj t_i]]

However, this approach has a number of problems. Firstly, there is no clear reason to trigger movement. A second problem is, strongly connected with the first one, the location of movement. If the constituent moves to a spec position, what triggers the movement? If it adjoins to somewhere, what triggers the adjunction, too. Therefore, it is more reasonable to assume that V and O move somewhere separately.

3.2 Word order correlations

In Greenberg's linguistics universals, there are number of cross-categorial correlations. For example, SOV languages tend to be postpositional, but language with VSO order are always prepositional. In terms of the head-complement parameter, Japanese exhibits a harmonious word order, wherein the order is cross-categorically head-final. In addition, Turkish has a head-final determiner, as illustrated in (5).

(5) Köfteyi   yiyor
     meatballs+the   he-is-eating

In fact, the definite article does not exist in Turkish. Instead, it appears as a suffix *-i* that is added to the end of a noun. It can be said that the D head follows NP in this case.

On the other hand, the head-complement order is not consistent in many languages. For example, an adjective usually precedes a noun (AN) in Germanic languages such as English and German although a noun should be followed by an adjective (NA) in head-initial languages within the Greenbergian framework (cf. Greenberg (1963)). Indeed, in Romance languages like Spanish, French, and Italian, it is the case that A follows N (e.g. *casa blanca* = *house white*).

Next consider Universal 6, as outlined below.

#6   All languages with dominant VSO order have SVO as an alternative or the only alternative basic order.

In the Minimalist framework, the strong preference between VSO and SVO is predicted since VSO is derived from SVO by the head movement of V over S to C, based on Celtic languages like Irish or Welsh.

(6) [$_{CP}$ SPEC Vi [$_{TP}$ S t_i [$_{vP}$ [$_{VP}$ t_i Obj]]

(7) *Thug   Aodh Rua a chaisleán do Mhaolmhuire.*
     give.PST Hugh Red his castle   to Maolmhuire

(Nolan 2021: 151)

It is worth mentioning that Amis does not obey Greenberg's (1963) linguistic universals 6. Liu (1999: 28) describes that actor-voice sentences in Amis can be either VSO or VOS, while

non-actor voice (or undergoer voice, UV) sentences can only be VOS[9].

(8) a. Mi.tilu   [ø-ci   aki] [t.u   fafui] i   lutuk anudafak.

      AV.hunt   Nom.PPN Aki Dat.CN   pig PREP   mountain tomorrow

   b. Mi.tilu   [t.u   fafuy] [ø-ci   aki] i   lutuk anudafak.

      AV.hunt   Dat.CN pig   Nom.PPN Aki PREP   mountain tomorrow

   'Aki will hunt pigs in the mountains tomorrow.'

In (8a) and (8b), the subject (*Aki*) and the object (*pigs*) are case-marked in the same manner. Therefore, this is not a case of SVO in Seediq.

3.3 head-complement variation in Seediq

In this section, I will explorer the internal order and structure of several phrases in order to see whether Greenberg's (1963) linguistic universals and Cinque's (2005) subsequent study hold for V-initial languages such as Seediq. This observation is based on the assumption that Seediq and Amis are head-initial given their VOS or VSO orders.

First of all, consider the internal order of DP in Seediq. Seediq has three main cases: nominative, oblique, and genitive. Tsukida provides (9) as an example of VOS. In this case, the verb is intransitive, and the second element has an oblique case.

(9) mawsa   karin.ku   ka.rubiq

    AV.FUT.go Hualien.OBL NOM.Rubiq

    'Rubiq will go to Hualien.'

                                   (Tsukida (2009: 143))

On the other hand, (10) is a transitive sentence with an oblique object preceding the nominative subject. In this point, the oblique case behaves similarly to the accusative case in English and Japanese.

(10) ('adi) gaga   s-em-pug   pat.as   ka.kumu.

    NEG DIST[10].PRG   AV-read   book.OBL NOM.Kumu

    'Kumu is (not) reading a book.'

                                     (Tsukida (2009: 276))

Next, (11a) and (11b) are ditransitive sentences. In both sentences, the verbs specify three

---

[9] PPN and CN stand for 'personal proper noun' and 'common noun,' respectively.

[10] DIST stands for 'distance,' with a deictic meaning.

arguments: nominative, genitive, and oblique. Judging from their meanings, nouns marked with genitive (*potato* and *paint*) function as direct objects, while nouns marked with oblique serve as indirect objects. Finally, nominative subjects occupy the final position in the sentence.

(11) a. m-egay   bu.na   leqi-'.an   ka.bubu
        AV-give sweet potato.GEN   child.OBL   NOM.mother
        'The mother gave sweet potato to the/a child.'
    b. ga   r.em.isuh   pin.ki   qenabil   ka.tama.
        PRG   AV.paint   paint.GEN   wall.OBL   NOM.father
        'Father is painting paint on the wall.'

<div align="right">(Tsukida (2012: 82))</div>

    c. [V + direct object + indirect object] + nominative subject

It is quite interesting that the genitive and oblique cases are affixed to the right of the noun (postpositional), while the nominative case is affixed to the left of the noun (prepositional).

The sentential subject follows even an adverbial element in (12).

(12) a. mwasa   Karin.ku   kusun   ka.rubiq
        AV.FUT.go   Hualien.OBL   tomorrow   NOM rubiq
        'Rubiq will go to Haulien tomorrow.'

<div align="right">(Tsukida (2012: 320))</div>

    b. [V + oblique object + adverb] + nominative subject

In this way, the sentential subject follows the second object and an adverb.

Next, consider the internal order of noun phrases. Cinque (2005) found that, with regard to Dem, Num, Adj, and N, the most frequent sequences are {Dem, Num, A, N} and {N, A, Num, Dem}. These orders represent underlying sequences for head-initial and head-final languages. However, it is well-known that Seediq does not have clear criteria to distinguish adjectives from verbs. Instead, as shown in (13), the noun '*stone*' modifies the preceding '*house*' refering to a 'stone house.'

(13) niqab sapah betunux.
    exist house stone
    'There is a stone house.'

I assume that this is a case of a lefthand-head compound. In other words, a modifier, equivalent to an adjective, follows a noun in Seediq. In fact, an adjectival verb modifies the preceding noun in (14). According to Tsukida (2009: 204), the opposite order makes a sentence ungrammatical.

(14) Kese'N.an　　[baki　　　]　[[ka se'diq]　　mpe.dawi]
　　　scold-GV2　　oldman.GEN　　NOM person　AV.FUT.idle
　　　'The/An old man scolds the idle person.'

　Let us now consider a complex NP structure. In (15), Num is placed at the initial position, and Dem, which follows the noun, appears in the final position in NP in (15b).

(15) a. me-'iyah　　　　　　　[ka　　　[teru kuyuh]]
　　　　AV.FUT.come　　　　　NOM　　three woman
　　b. [laqi　masaw]　　　　　[ka　　　[deha　　laqi　　niyi]]
　　　　child　Masaw.Gen　　NOM　　two　　child　　this
　　　'These two children are Masaw's child.'
(16) a. [case [NUM — N]]
　　b. [case [NUM — N —Dem]]

(16b) illustrate the internal structure of NP: [NUM　— N　—Dem]. However, Cinque (2005) reported that this sequence, similar to (17a) and (17b), was not attested[11].

(17) a. Num N Dem A　　　　　　　(= (6g), Cinque (2005: 319))
　　b. Num N A Dem　　　　　　　(= (6s)), Cinque (2005: 320)

　In Seediq, D is postnominal. In (18), there seems to be a lack of a verb. Like Semitic languages, the copula '*be*' does not appear in the present tense in Seediq. Thus, this sentence is predicate-initial, as usual. On the other hand, Amis is much harmonious than Seediq, in that only Gen follows N, while Dem, Num, and A are prenominal.

(18) patas senaw=mu　　　　　　　ka.niyi
　　　book husband.Gen=1s.GEN　　NOM.this
　　　'This is my husband's book.'

---

[11] See Appendix for the complete list of the combinations.

(19) wada   me.huqil   [[ka.wana   kiNal=bi   [laqi=na]]
     PST   AV-die   NOM only   one=really   child=3s.GEN
     'His/Her only child died.'

Another case involves an NP with multiple complements. As the English translation indicates, example (20) appears to be a sentence, but, in fact, it is a noun phrase.

(20) [haNan 'uwa   [p.en-we'la nihun]]   niyi]] 'u,   Nawmi.
     name girl.GEN   PRF-from   Japan.OBL   this   CNJ   Naomi
     'The name of this girl from Japan is Naomi.'

<div align="right">(Tsukida 2009: 302))</div>

In (20), the subject is '*name*,' which is the head of NP and the subject of the sentence. The internal structure of the subject NP is so complex; the genitive NP modified by PP also selects *name* as its complement. Furthermore, the demonstrative in the final position of the whole NP modifies the initial N, *name*. Look at (21) below.

(21) [NP name [DP girl's [PP from Japan]] this] = [NP N [DP N [PP P NP]] Dem]

In this structure, P precedes NP, the genitive NP selects the NP, which is then modified by the demonstrative *this*.

3.4. possible and impossible orders

We have observed several orders of syntactic items, which is summarized below (=(22)). First, consider the arrangement of nominal elements in noun phrases. If we assume that a modifier is equivalent to an adjective, Seediq has three potential arrangements in NP.

(22) a. [NUM N DEM]   (NUM>N, N>DEM)
     b. [N Modifier]
(23) a. [Num N Dem A]
     b. [A Num N Dem]
     c. [Num N A Dem]   !!!

(23a) and (23b) are not attested in Cinque (2005), and (23c) is documented in few languages, including some Austronesian languages such as Indonesian, Jacaltec, and so on (Cinque (2005: 320)). It follows that Seediq also has this Num-N-A-Dem order. How is this order derived? If

the canonical order is [Dem Num A N] or [N A Num Dem], several head or phrasal movements are required to derive (23c). For example, one of the simplest ways to derive [Num N A Dem] from [Dem Num A N] is a partial movement of [Num A N] out of the entire phrase. However, this is almost impossible because [Num A N] cannot form a constituent. Therefore, each head must move leftward/upward independently. Alternatively, (23c) can be achieved through the movement of Num to the leftmost position of [N A Num Dem] as long as certain reasons trigger the movement.

   Next, turn to the configuration of prepositional phrases. Almost all the cases are consistent with the head-initial order. (24a) is a case of a preposition and (24c) is prenominal; therefore, they are harmonious. On the other hand, only in (24b), the case marker is postnominal with a head-final configuration. There is no clear reason of this discrepancy.

   (24) a. [P NP]
        b. [NOM N]
        c. [N OBL], [N GEN]

There are a number of problems with head-complement orders. In particular, the underlying structure of VOS and the process of deriving the VOS order has not been fully explained yet.

4. Summary

This paper has explored the harmonious/disharmonious head-complement order of Seediq, with a basic word order of VOS. On the assumption that VOS is head-initial, it can be expected that head-complement orders in Seediq should also be head-initial. However, the data we collected are not satisfactory enough to demonstrate whether Seediq word orders are consistent with Greenberg's (1965) generalization. This is partly because the number of VOS languages is small, and there have been few theoretical studies on VOS. Thus, further investigations are required to explain the consistency of head-complement orders are in minor languages.

Appendix: 24 combinations of four elements in NP (Cinque (2005))

a	OK	Dem Num A N	very many languages
b	OK	Dem Num N A	many languages
c	OK	Dem N Num A	very few languages
d	OK	N Dem Num A	few languages
e	NG	Num Dem A N	ø
f	NG	Num Dem N A	ø
g	NG	Num N Dem A	ø
h	NG	N Num Dem A	ø
i	NG	A Dem Num N	ø
j	NG	A Dem N Num	ø
k	OK	A N Dem Num	very few languages
l	OK	N A Dem Num	few languages
m	NG	Dem A Num N	ø
n	OK	Dem A N Num	very few languages
o	OK	Dem N A Num	many languages
p	OK	N Dem A Num	very few languages
q	NG	Num A Dem N	ø
r	OK	Num A N Dem	very few languages
s	OK	Num N A Dem	few languages
t	OK	N Num A Dem	few languages
u	NG	A Num Dem N	ø
v	NG	A Num N Dem	ø
w	OK	A N Num Dem	very few languages
x	OK	N A Num Dem	very many languages

References

Aldridge, Edith. 2019. Labeling and verb-initial word order in Seediq. *Journal of East Asian Linguistics 28*: 359-394.

Chomsky, Noam., Ángel J. Gallego, and Dennis Ott. 2017. Generative Grammar and the Faculty of Language: Insights, Questions, and Challenges. *Catalan Journal of Linguistics Special Issue*: 229-261. (https://doi.org/10.5565/rev/catjl.288)

Cinque, Guglielmo. 2005. Deriving Greenberg's Universal 20 and Its Exceptions. *Linguistic Inquiry 36*: 315-332.

Cinque, Guglielmo. 2018. Some notes on "meaningless movement." ms. Università di Ca' Foscari.

Gell-Mann, Murray and Merritt Ruhlen. 2011. The origin and evolution of word order. *PNAS 108 (42)*: 17290-17295. (https://doi.org/10.1073/pnas.1113716108)

Greenberg, Joseph. H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg (Ed.), *Universals of Language*: 73-113. Cambridge, MA: MIT Press.

Karlsson, Anastasia and Holmer, Arthur. 2011. Interaction between word order, information structure and intonation in Puyuma. In M. Endo (Ed.), *Linguistic Circle for the Study of Eastern Eurasian Languages*: 28-38.

Kayne, Richard S. 1994. *The Antisymmetry of Syntax*. Cambridge, MA: MIT Press.

Kayne, Richard S. 2018. The Place of Linear Order in the Language Faculty. ms. Venice University.

Liu, Dorinda Tsai-hsiu. 1999. *Cleft Constructions in Amis*. Taipei: National Taiwan University MA Thesis.

Nolan, Brian. 2021. Word Order Alignment in Three-Argument Constructions of Irish. *Roczniki Humanistyczne 69 (11)*: 149-169.

Ono, Hajime, Jungho, Kim, Manami, Sato, Apay, Ai-yu, Tang, and Masatoshi Koizumi. 2020. Syntax and processing in Seediq: a behavioral study. *Journal of East Asian Linguistics 29*: 237-258. (https://doi.org/10.1007/s10831-020-09207-7)

Tsukida, Naomi. 2009. *Sedekku-go-no Bunpou* (Grammar of Seediq). Ph.D. Dissertation, the University of Tokyo.

Tsukida, Naomi. 2012. Voice and Conveyance Voice of Seediq. *SENRI ETHNOLOGICAL STUDIES 77*: 77–95.

Vennemann, Theo. 1973. Explanation in syntax. In J.P. Kimball (ed.), *Syntax and Semantics 2*: 1-50. New York: Seminar Press.

URL

The World Atlas of Language Structures

http://wals.info

Accessed on December 27, 2023