

# DiACL – Diachronic Atlas of Comparative Linguistics Online. Description of Dataset DiACL/ Typology/ Austronesia

**Author:** Arthur Holmer (Lund University)

**Date of publication:** 2017-03-10

**Contributors** (dataset design, feeding, language control/ contribution, database construction, programming, advice): Arthur Holmer, Niklas Johansson, Sandra Cronhamn, Filip Larsson.

**Reference, database:** Carling, Gerd (ed.) 2017. Diachronic Atlas of Comparative Linguistics Online. Lund University. Accessed on: z

**URL:** <https://diACL.ht.lu.se/>

**Reference, dataset:** Holmer, Arthur, Niklas Johansson, Sandra Cronhamn, Filip Larsson 2017. Diachronic Atlas of Comparative Linguistics Online. Dataset DiACL/ Typology/ Austronesia. Accessed on: z

**Reference, document:** Holmer, Arthur (2017). DiACL – Diachronic Atlas of Comparative Linguistics Online. Description of dataset Typology/ Austronesia. Lund University, Centre for Languages and Literature.

## Table of Contents

§1. Introduction .....	1
§2. Word Order .....	2
§3. Relativization .....	5
§4. Subject properties .....	5
§5. Verb inflection on other elements .....	6
§6. Content interrogatives .....	7
§7. Placement of particles .....	8
§8. Specifically Formosan properties: connegativity and prefix concord .....	9
§9. Summary .....	10
References .....	10

## §1. Introduction

One major weakness of traditional typological classification is that it treats languages and linguistic phenomena as monolithic entities. However, the grammatical properties of any language are usually much more complex, and in any such classification, information is invariably lost. Further, the exact nature of what is lost depends to a great extent on the subjective judgment of the researcher contributing the data. To avoid this double problem, the DiACL database capitalizes on the atomic properties underlying a given classification, and addresses them directly.

For example, while a language may be monolithically classified as having an ergative, accusative or active case pattern, case alignment is the result of a synthesis of various atomic factors, and can be readily reduced to them. Thus, an accusative case alignment implies that A, Sa and So share one case form, while O carries a different case form. An active case alignment implies that A and Sa share one case form, while So and O share another. An ergative case alignment implies that Sa, So and O share one case form, while A bears a separate case form. These atomic properties can vary independently of one another yielding various types of active or tripartite systems, which are partially similar and partially distinct. This strategy carries over to other phenomena. Specifically, in the Austronesian survey, the clitic status of pronominal arguments is a function of the difference between word order with full NP arguments and word order with pronominal arguments, and the verbal (head) status of adverbials and negations is the function of the difference between clauses containing such elements and clauses lacking them.

Further, in keeping with the general idea of DiACL, mutually exclusive facts are listed as separate datapoints. The purpose of this is to adequately capture variation within the language. Without this feature of the database, variation would either lead to a label such as *Free word order*, or *No uniform word order data available*, or even to the erroneous classification of a language based on the subjectively perceived relative frequency of different patterns on the part of the coder. Even in cases where there is a clear preference, the system applied here allows us to capture the fact that other alternatives are also attested (while others are not attested), yielding a finer degree of granularity.

In the database, the data sets are ordered in such a way as to allow a progression in the coder's understanding of the language while mining for data in published reference grammars. Thus, in the Austronesian context, which displays two major argument structure categories (illustrated by Actor Voice and Non-Actor Voice respectively), word order must be defined independently for both. Given the subject-orientation of Austronesian relativization, the most straightforward method to find an example of each is to compare the verbal morphology in object relativization (*the book that Mary bought*) and subject relativization (*the girl that bought the book*). Therefore, questions on relativization, as well as applicatives (sets 13–15), which are thematically related, are interspersed between questions dealing with word order (sets 1–12, 17–27). In the discussion we will instead deal with the issues thematically, regardless of the order in which they appear in the database.

## §2. Word Order

Word order is listed separately for transitive verbs and for intransitive verbs (sets 1, 2), to capture any instances of ergative word order (i.e. where S and O share a typical position not shared by A). Further, in many Austronesian verb-initial languages, subject topicalization is a very frequent phenomenon, so frequent, in fact, that it is quite possible that a cursory overview of the word order might erroneously view SVO derived by topicalization as the basic order. Finally, languages which do display SVO word order may well have derived this from overgeneralization of the subject topicalization construction. Apparent subject-initial constructions do not only involve topicalization, but can also be instances of clefting, which is usually recognizable by the fact that the following VP is marked morphologically as a subject.

- 1     a)    Rseno      kiya      wada      m-angan      budi                    *Tgdaya Seediq*  
           man        that       PST       ActV-take    arrow  
           ‘That man took an arrow.’

- b) Yaku ka m-n-imah sino. *Tgdaya Seediq*  
 1sN NOM ActV-PST-drink wine  
 ‘I’m the one who drank wine.’

For these reasons, sets (1–3) through include datapoints dealing with subject fronting of various kinds, overt clefting and overt topicalization.

Austronesian languages of the Formosan/Philippine type display a multipolar voice system. The exact number of voices varies from language to language, with the original four-voice system reconstructed to Proto-Austronesian (Wolff 1973, Starosta, Pawley & Reid 1982) being reflected directly in e.g. Seediq and Tagalog, while Malagasy has reduced this to a three-voice system for many verbs. In Tagalog, the same proposition can be realized in four different ways, with varying forms of the verb, depending on which of the arguments of the verb is the theme of the proposition (Schachter 1976: 494–495).

- 2 a) Mag-alis ang babae ng bigas sa sako para sa bata.  
 ActV-take NOM woman OBL rice LOC bag for child
- b) Aalis-in ng babae sa sako para sa bata ang bigas.  
 take-PatV GEN woman LOC bag for child NOM rice
- c) Aalis-an ng babae ng bigas para sa bata ang sako.  
 take-LocV GEN woman OBL rice for child NOM bag
- d) Ipag-aalis ng babae ng bigas sa sako ang bata.  
 BenV-take GEN woman OBL rice LOC bag NOM child
- ‘The woman will take the rice out of the bag for the child.’

The exact nature of this system is not relevant for the purposes here; one account (Chang 2015) is that they are to be seen as applicative morphology acting on a syntactically ergative system. What is important, though, is that the choice of voice generally (but not always) affects word order. Thus, in e.g. Tsou (3) and Malagasy, the nominative argument (grammatical "subject") is generally clause-final, regardless on whether it refers to the AGT or PAT.

- 3) mo mo-si ta pangka to emi ’o amo.  
 INTR.RL INTR-put OBL table OBL wine ABS father  
 ‘Father put wine on the table.’ (Chang 2015:3)

In contrast, in Tagalog, while there is also a tendency for the nominative argument to be clause-final, this is sometimes overridden by a preference for the early placement of an AGT. Thus, in Actor Voice,

these two principles compete and word order varies between VSO and VOS, whereas this does not occur in Non-Actor Voice (4c, d).

- 4 a). s<um>ulat      ng=liham      si=Juan      **VOS**  
 <ActV>write      OBL=letter      NOM=Juan  
 ‘Juan wrote a letter.’ (Billings 2005:307)
- b). s<um>ulat      si=Juan      ng=liham      **VSO**  
 <ActV>write      NOM=Juan      OBL=letter  
 ‘Juan wrote a letter.’ (ibid.)
- c). ?s<in>ulat      ang=liham      ni=Juan      **?VSA**  
 <PatV>write      NOM=letter      OBL=Juan  
 ‘Juan wrote a letter.’ (Kroeger 1993:111)
- d). s<in>ulat      ni=Juan      ang=liham      **VAS**  
 <PatV>write      OBL=Juan      NOM=letter  
 ‘Juan wrote a letter.’ (ibid.)

For this reason, the first few sets of questions on the word order of transitive verbs include a specific requirement that the form of the verb be the same as for an intransitive verb, i.e. in Actor Voice (sets 3–12). This is followed by a set of questions dealing with non-Actor Voice (sets 17–27). Addressing word order in this fashion allows us to capture the very specific word order patterns in several Austronesian languages, without having to determine a priori which of the two constructions Actor Voice and Non-Actor Voice is to be selected, by virtue of being prototypically transitive, as the only construction to be used to exemplify word order.

Many Austronesian languages also have clitic pronouns with a linear distribution which is different to that of full NPs. Therefore the word order datapoints for various clause types includes three types of argument configuration: a) with both arguments realized as full NPs (sets 3–7, and 18–22); b) with one argument realized as a full NP, and one argument realized as a clitic pronoun (sets 8–12); and c) with both arguments realized as clitic pronouns (sets 23–27).

Word order is also affected by various sets of preverbal elements. In Bunun, which displays VSO order in Actor Voice, the position of the verb can be occupied by the negation (5). Further, in all Formosan languages, most adverbial meanings (including all adverbials of manner and frequency) are realized as inflected verbs, occupying verbal position, but also behaving in other ways as verbs, e.g. bearing finite verbal inflection. Directional coverbs in serial verb constructions, which sometimes also serve as periphrastic tense markers, also occupy preverbal position and interact with word order in various ways.

- 5) Ni ca Alang ma-ludaq tama-nak. *Takituduh Bunun*  
 NEG NOM Alang ActV-beat father-1sGEN  
 ‘Alang did not beat my father.’

Finally, subordinators, in particular the conditional subordinator *if*, sometimes also affect word order in a similar manner. Thus, in the VOS language Seediq, the subordinator attracts a clitic pronoun to a position preceding the verb, while this does not happen in Bunun.

For these reasons, the data sets target the independent variation of voice, word order with full NP arguments and clitic pronouns, and various types of verb-like elements, or other elements which display similar behaviour. Datapoint (12d), which targets word order *A-if-VO* might seem to be superfluous in an Austronesian context, and just a replica of datapoints (9d), (10d) and (11d), but is deliberately included to track possible influence from Taiwanese Hokkien, the majority language of Taiwan, which has exactly this word order pattern.

The result is a cross-linguistically commensurable description of word order which is considerably more fine-grained than any which has hitherto been reduced to a feature matrix. In addition, in each feature being explicitly defined, it avoids the problem of lumping languages into monolithic classifications such as "VSO" or "VOS".

### §3. Relativization

Over and above the use for identifying the Actor Voice and Non-Actor Voice word order patterns, relativization is an ideal mechanism for capturing different variants of the voice system in Austronesian, given the fact that Austronesian relativization tends to be subject-oriented (set 13 is designed to check this claim for each language). Set 14 addresses the typical Western Austronesian multipolar voice system to distinguish it from e.g. the Malay two-voice system, without specifically addressing the issue of exactly how many different voice forms are proposed for the language (an issue where there is some debate, since scholars differ on which formatives to include as voice categories of voice). Further, to make sure that the distinction is voice rather than applicativity, a separate datapoint on applicatives is included (set 15).

Relativization constructions in Austronesian are sometimes treated as a type of nominalization construction, which derive a filler-gap dependency pragmatically rather than syntactically. To capture this empirically, set (17) tests for the existence of gapless relatives, which would be expected under such an analysis. Needless to say, this set is very hard to fill from reference grammars alone, and is therefore not very populated. Set (28) tests directly for the behaviour of action nominalizations, with respect both to the case forms of the arguments and to the form of the verb. A related issue is whether the agent in a Non-Actor Voice construction is realized in genitive or in another case form (set 16). This touches directly on the ergativity analysis of Austronesian (e.g. Aldridge 2004), for whom the genitive case of the agent is an instance of ergative case.

### §4. Subject properties

Since Schachter (1976), the issue of subject properties in Austronesian languages has been a schoolbook example of how subject properties can vary across languages. Over and above the question of case marking in Austronesian Non-Actor Voice, three of the classical Schachter subject property tests are included: reflexives (set 38), conjunction reduction (set 39) and floating quantifiers (set 40). As the result of our current understanding of Austronesian morphosyntax, several possible values are available,

including allowing for the possibility that reflexives (as in Takituduh Bunun, 6a) or quantifiers (as in Tsou, 6b, c) can be realized morphosyntactically as verbs.

- 6 a) anakanak-ak      sadu      ha'an      tidanuman  
 REFL-1sN      see      LOC      mirror  
 'I see myself in the mirror.'
- b) Te-hin'i      aciH-1      cmi'ho.  
 IRR-3p      all-ActV      come.ActV  
 'They will all come.'  
*Tsou* (Chang 2002:330)
- c) Os'ocu      aciH-a      an-a      'e      kamae.  
 NAV-1sG-ASP      all-NAV      eat-PatV      NOM      guava  
 'All the guavas have been eaten up by me.'  
 (ibid.)

## §5. Verb inflection on other elements

In Formosan languages, the first primary branches of Austronesian, manner adverbs are uniformly morphologically realized as verbs, both as concerns syntactic distribution and inflectional and derivational morphology (7a, b; data from Tgdaya Seediq). However, it appears that in several languages (i.e. those languages for which such data is currently available, including Tsou and Tgdaya Seediq) the voice realization of such preverbs is not multipolar, but rather only binary (Actor Voice / Non-Actor Voice). In a construction with such a preverb, the lexical verb can either bear the full (multipolar) voice morphology of the clause (as is the case in Tsou, where the lexical verb and the adverbial verb essentially agree in voice, 7c) or default Actor-Voice morphology (as is the case in other languages). These differences are captured by datapoints (29a–c). For most languages this data is not yet readily available, but will be filled in successively. It is an interesting dimension, given that morphological agreement across verbal elements in Tsou seems to be a similar kind of phenomenon to prefix concord (see the corresponding section). Furthermore, in Malagasy, prepositions are sometimes marked for (past) tense, although the formative is not the same as for the verb (7c, d; cited from Pearson 2005). This is another intriguing fact concerning agreement across categories. It is targeted by set (42), initially in an attempt to trace similar phenomena in the closest relatives of Malagasy, e.g. Ma'anyan.

- 7 a) M<n>hmet-an-mu      m-imah      sino      kiya.  
 <PST>at.will-LocV-1sGEN      ActV-drink      wine      that  
 'I drank that wine with no thought about the consequences.'
- b) ini-mu      p-hmet-i      m-imah      sino      (ka)      laqi=mu  
 NEG-1sGEN      CAUS-at.will-PatV.CNG      ActV-drink      wine      NOM      child-1sGEN



## §7. Placement of particles

The placement polar interrogative particle varies greatly in Austronesian: in Seediq it is clause-initial (9a), in Atayal it is clause-final (9b), in Tagalog it is a second-position clitic (9c) and in Malagasy it follows the verb phrase, immediately preceding the subject (9d). Set (33) covers these options, and opens for other possibilities as well. In Tagalog, there are many other clause-level elements which are also realized as second-position clitics, including adverbials such as *na* ‘already’ (9e), *pa* ‘still’ (9f), *din/rin* ‘also’ (9g) as well as the evidential *daw/raw* ‘apparently’ (9h). At first blush, comparing Tagalog with some Formosan languages, there seems to be an interesting connection between the realization of these particles and clause-level word order (realization as second-position clitics seems to correlate with VSO order, while other placements correlate with VOS order). To test whether this connection can be confirmed over a larger set of languages, these particles are addressed in datapoints (33) through (37).

- 9 a) **ye-su**            m-n-ekan            hlama            kiya?            *Seediq*  
 Q-2sNOM    ActV-PST-eat    steamed.rice    that  
 ‘Did you eat that steamed rice snack?’
- b.    **kia'**    'i'            ma-qilaap    'i'            **quw?**            *Atayal (Mayrinax)*  
 Asp    Link    ActV-sleep    NOM    mother    Q            (Huang 1995:176)  
 ‘Is mother sleeping?’
- c)    **B<um>ili**            **ba** ng            karne            ang            Nanay?            *Tagalog*  
 <ActV.PST>buy    Q    OBJ    meat            NOM    mother            S&O: 502  
 Did mother buy some meat?
- d)    Nanome    vola            an-dRabe    **ve**            ianao?            *Malagasy*  
 gave            money            ACC-Rabe    Q            you            (Keenan (1976:252))  
 ‘Did you give Rabe money?’
- e)    Hindi    **na**            ako            pu-punta            roon.            *Tagalog*  
 NEG    PRF    1s            RED-go            there            S&O: 418  
 ‘I won’t go there anymore’ (i.e. It is already the case that [I do not go there]).
- f)    Hindi    **pa**            siya            um-a-alis            *Tagalog*  
 NEG    yet            3s            ActV-RED-leave            S&O: 419  
 ‘S/he hasn’t left yet.’

- g) Um-a-awit **din** si Carmen *Tagalog*  
 ActV-RED-sing also NOM Carmen S&O: 422  
 ‘Carmen is singing, too.’
- h) Mabuti **raw** ang ani. *Tagalog*  
 good QUOT NOM harvest S&O: 423  
 ‘They say that the harvest is good.’

### §8. Specifically Formosan properties: connegativity and prefix concord

We have selected two properties which have a contiguous distribution across various primary sub-branches in Taiwan: connegativity (set 30) and prefix concord (set 41). These are deliberately chosen to capture what appears to be either areal effects or possibly archaisms which potentially may have reflexes in other Austronesian languages.

Prefix concord, which is found in Siraya (†), Tsou and Bunun in south-central Taiwan (Nojima 1996), is a phenomenon whereby auxiliary-like elements such as adverbial verbs agree in semantic category with the lexical verb (10a, b). Connegativity, which is found in the Atayalic languages and Amis (as well as in some Baltic-Finnic languages like Finnish and Saami), implies that the first verb (or auxiliary, adverbial verb etc.) which follows after a sentence negation must be realized in a form which is morphologically identical to the imperative (10c, d).

- 10 a) Pit-hiav-ak pit-dias qaising.  
 COOK-quick-1sN COOK-cook rice  
 ‘I cook rice quickly.’
- b) Mis-utmag mis-busuk. *Bunun*  
 ADVERSE-careless ADVERSE-drunk (Nojima 1996: 17)  
 ‘(He) carelessly became drunk.’
- c) Ini-mu mah-i ka sino.  
 NEG-1sGEN drink-PatV.CONNEG NOM wine  
 ‘I don’t / didn’t drink the wine.’
- d) Mah-i ka sino nii!  
 drink-PatV.IMP NOM wine this  
 ‘Drink this wine!’

## §9. Summary

Naturally, any selection of features is a subjective choice. The guiding principle behind the features we have selected for the Austronesian data set has been in particular to address features where we know that there is a great variation within Austronesian (e.g. word order and wh-constructions), features which deal with the issue of subject properties and ergativity, which is an ongoing debate among Austronesianists at present and a small sample of features which appear to have an areal rather than genealogical distribution.

## References

- Aldridge, Edith. 2004. *Ergativity and word order in Austronesian languages*. PhD thesis, Cornell University.
- Billings, Loren. 2005. Ordering clitics and postverbal R-expressions in Tagalog. A unified analysis? In Andrew Carnie, Heidi Harley & Sheila Ann Dooley. *Verb first. On the syntax of verb-initial languages*. Amsterdam: John Benjamins
- Chang, Henry Y.-L. 2002. Distributivity, plurality and reduplication in Tsou. *Tsing Hua Journal of Chinese Studies* 32(2): 327–355
- Chang, Henry Y.-L. 2015. Extractions in Tsou causative applicatives. *Lingua Sinica*. 1: 5 (doi:10.1186/s40655-015-0003-7)
- Huang, Lillian. 1995. *A study of Mayrinax syntax*. Taipei: Crane Publishing
- Keenan, Edward. 1976. Remarkable subjects in Malagasy. In Li, Charles (ed). *Subject and topic*. New York: Academic Press; 247-301
- Kroeger, Paul. 1993. *Phrase structure and grammatical relations in Tagalog*. Stanford: CSLI
- Nojima, Motoyasu. 1996. Lexical prefixes of Bunun verbs. *Journal of the Linguistic Society of Japan* 110: 1-27.
- Pearson, Matt. 2005. Tense (mis)matches between verbs and obliques in Malagasy. In Heinz & Ntelitheos (eds.). *UCLA Working Papers in Linguistics, no. 12, September 2005. Proceedings of AFLA XII*, 369–382
- Schachter, Paul. 1976. The Subject in Philippine Languages: Topic, Actor, Actor-Topic, or None of the Above. In Li, Charles (ed.). *Subject and Topic*. New York: Academic Press.
- Starosta, Stanley, Andrew Pawley & Laurence Reid. 1982. The evolution of focus in Austronesian. In Amran, Halim, Lois Carrington & S.A. Wurm (eds.). *Papers from the Third International Conference on Austronesian Linguistics, vol. 2*. Canberra: Pacific Linguistics. 145–170.
- Szakos, Jozsef. 1994. *Die Sprache der Cou. Untersuchungen zur Synchronie einer austronesischen Sprache auf Taiwan*. PhD dissertation, Friedrich-Wilhelms-Universität, Bonn.
- Wolff, J. U. 1973. Verbal inflection in Proto-Austronesian. In Gonzalez, A.B. (eds.) *Parangal kay Cecilio Lopez: Essays in honor of Cecilio Lopez on his seventy-fifth birthday*. LSP Special Monograph Issue No. 4. Quezon City: Linguistic Society of the Philippines: 71-91.