

Word order and disambiguation in Pangasinan

There is a well-known typological correlation between word order flexibility and overt case and/or agreement (Sinnemäki, 2008; Fedzechkina et al., 2017; a.o.). This correlation receives a straightforward functionalist explanation: the mapping between arguments and verbs (e.g. theta role mapping) should be relatively unambiguous, either by rigid word order or case/agreement. The “choice” between these options is often conceptualized as a *language-level* parameter.

In this talk, we report on clause structure and word order in Pangasinan (Austronesian; northern Philippines), based on original elicitation work. Pangasinan has case marking and some agreement (described below) and generally free postverbal word order, as expected typologically and based on the behaviors of other Philippine languages. However, in a particular, limited configuration where case and agreement fail to disambiguate between core arguments, their word order becomes rigid. We argue that an individual language’s grammar may formally restrict operations such as scrambling in cases that would lead to ambiguity.

Pangasinan clause types: Pangasinan grammar has a Philippine-type voice system, similar to Tagalog, but with a relatively rich inventory of case markers (Benton, 1971). In each clause, one core argument — the “pivot” — has a privileged status, for example in being nominative-marked and allowing \bar{A} -extraction. This correlates with the form of the verb: the agent is the pivot in Actor Voice (1a), the theme is the pivot in Patient Voice (1b). In these clause types, postverbal word order is free, just as in other Philippine languages: the two DP arguments can be in either order, and other adjuncts and oblique arguments may be interspersed in any position.

- (1) a. Actor Voice: AV-V { NOM=DP_{agent}, GEN=DP_{theme}, ... } ← free word order
b. Patient Voice: PV-V { GEN=DP_{agent}, NOM=DP_{theme}, ... } ← free word order

We also discuss a third clause type, which involves a Patient Voice verb with a genitive clitic pronoun doubling the agent, and where both the agent and theme are nominative.

- (2) Double NOM: PV-V=*pro*_i { NOM=DP_{agent,i}, NOM=DP_{theme}, ... } ← *mostly* free word order
We descriptively refer to this structure as the “double nominative” clause type.

Word order in the double nominative: When the agent and theme differ in ϕ -features, word order is free. (3) has a third-plural agent and third-singular theme. Because the clitic pronoun =*da* unambiguously doubles the agent, we know that the plural DP ‘the boys’ is the agent.

- (3) Luluto-en =da { ra-**may** lakin ugaw **may** sira / **may** sira ira-**may** lakin ugaw }.
PROG.COOK-PV GEN.3PL PL-NOM male child NOM fish NOM fish PL-NOM male child
‘The boys are cooking the fish.’

In third-singular/third-singular and third-plural/third-plural combinations with two *may* nominatives, postverbal word order is restricted: the agent must precede the theme. (Note that other material can intervene between the verb and arguments or between the arguments.)

- (4) P<in>uniti =to **may** lakin ugaw **may** bien ugaw.
PV-hit GEN.3SG NOM male child NOM female child
‘The boy hit the girl.’ (agent < theme) / *‘The girl hit the boy.’ (*theme < agent)

This rigidity cannot be overridden by semantic factors such as animacy or world knowledge. For example, (5) is judged as degraded as it can only describe an extremely unlikely scenario:

(5) Luluto-en =to **may** sira **may** lakin ugaw.

PROG.COOK-PV GEN.3SG NOM fish NOM male child

#‘The fish is cooking the boy.’ (ag < th) / *‘The boy is cooking the fish.’ (*th < ag)

This argues against the rigidity effect as the reflection of a default parsing strategy that assumes agent < theme order just in cases of ambiguity.

Rigidity is also not simply a matter of ϕ -feature non-distinctness. Pangasinan has two nominative case forms for common nouns: *may* and *su*. In (1a, 1b), the pivot can be marked with *may* or *su*. But in double nominative clauses (2), *su* is only available for the pivot theme. When *su* is used for the pivot, then, this too disambiguates, allowing for word order freedom:

(6) P<in>uniti =to { **may** lakin ugaw **su** bien ugaw / **su** bien ugaw **may** lakin ugaw }.

PV-hit GEN.3SG NOM male child NOM fem child NOM fem. child NOM male child

‘The boy hit the girl.’ cf (4)

Proposal: A background theory of Philippine voice: Our proposal follows existing work on the syntax of Philippine voice system languages (e.g. Aldridge, 2004; Rackowski and Richards, 2005). The “pivot” is the highest DP in the vP phase. In Patient Voice (7a), the theme covertly moves to an outer specifier of vP . In Actor Voice (7b), nothing moves around the agent.

(7) a. PV: [vP DP_{theme,i} [DP_{agent} [v [... V t_i ... b. AV: [vP DP_{agent} [v [... V (DP_{theme}) ...

Caseless DPs inside vP appear with default genitive case (Erlewine, Levin, and Van Urk 2020). Probes on T: On this basic theory, T Agrees with the closest DP to assign nominative case. We adopt this probe (i), and introduce a second, optional probe (ii). (8) gives the probes on T:

(8) i. obligatory; targets the closest DP and assigns it nominative (*su* or *may*).

ii. optional, probes second; targets the *next* closest DP and (a) copies its ϕ -features, to be realized as ϕ -agreement, and (b) assigns it restricted nominative case (*may*).

Probe (i) necessarily targets the pivot, the highest DP in vP , explaining the target’s appearance with *su* or *may*. Probe (ii) optionally probes after (i), targeting the agent; see (7a). In AV clauses (7b), there is only one DP at the vP phase edge, which receives nominative (i). There is no effect of optional probing by (ii), explaining the lack of a double nominative construction in AV.

Linearizing vP : Following Fowlie 2013 on Tagalog, we take the vP phase in Philippine languages to freely generate all possible word orders. (V moves out of vP .) Specifically when two DPs in vP are formally ambiguous — the same ϕ -features, both nominative case, with the same nominative case marker allomorph chosen — Kayne’s LCA is referenced to introduce a higher-ranked linearization statement: agent < theme, because the agent c-commands the theme’s overt position. (Movement of the theme pivot in (7a) is designated as covert movement.)

Selected references: Aldridge 2004 Cornell PhD • Benton 1971 *Pangasinan reference grammar* • Erlewine, Levin, Van Urk “The typology of nominal licensing in Austronesian voice system languages” • Fowlie 2013 “Multiple multiple spell out” • Rackowski & Richards 2005 “Phase edge and extraction” • Sinnemäki 2008 “Complexity trade-offs in core argument marking”