PROCESSING PHILIPPINE AND REVERSE-PHILIPPINE CASE ALIGNMENT*

Steven Foley
University of Georgia
srfoley@uga.edu

In morphosyntactic typology, the verb-initial Western Austronesian languages are well known for their complex 'Philippine Alignment' system for mapping grammatical/thematic roles onto case categories. The verb-final South Caucasian family also has complex case systems, which can be viewed as sort of 'Reverse Philippine Alignment'. This paper makes precise the connection between these typologically quite distinct grammars, in descriptive and information-theoretic terms. Synthesizing previous sentence-processing research on Tagalog and Georgian, I show that comparison between these systems has much potential to shed light on interactions between verb-initiality, verb-finality, and case marking as cues for the real-time comprehension of grammatical roles.

1. Introduction

Sentences are processed rapidly, incrementally, and actively (e.g. Ferreira & Qiu 2021). Key to understanding any sentence is identifying the event conveyed, and the thematic/grammatical roles of the event participants — that is, processing who did what to whom. The morphosyntactic encoding of grammatical relations varies dramatically across languages, and work in comparative sentence processing has made major strides understanding the real-time psycholinguistic consequences of this variation (e.g. MacWhinney & Bates 1989, Bornkessel-Schlesewsky & Schlesewsky 2009a,b). In language after language, it is clear that comprehenders attend to case and word order cues in order to make active predictions about a sentence's agenthood- and patienthood-relations. However, just a small proportion of attested morphosyntactic patterns have been considered in psycholinguistic research, either theoretical or empirical. The goal of this paper is to highlight a previously unexplored typological connection — between the Austronesian and South Caucasian language families — that might inspire comparative sentence-processing work to fruitfully advance our understanding of how grammatical roles are processed in real time.

To set the stage, consider some typical sentences of Tagalog (1). This language, like many Western Austronesian languages, encodes grammatical relations through a *Philippine Alignment* system (a.k.a. a Western Austronesian Voice system; Chen & McDonnell 2019). In a monotransitive clause, agents and patients are not uniquely associated with any particular linear position or morphological case category. Rather, a cue on the clause-initial verb — the 'voice' morpheme — determines which role is marked by which case. For instance, if the verb is in the 'agent voice' (1a), the event's agent will correspond to the noun marked by the particle *ang* 'ALPHA', while the patient will be marked by *ng* 'BETA'. But, if the verb is in the 'patient voice'

* Thank you to Norielle Adricula, Rajesh Bhatt, Ivan Bondoc, Rowena Garcia, Daniel Kaufman, Jed Pizarro-Guevara, and Matt Wagers, for insightful comments and discussion. I am very grateful for the organizers of AFLA 31 for organizing a psycholinguistics session that gave a Caucasianist like myself the chance to crosspollinate with a very friendly crowd of Austronesianists.

(1b), the case–role mapping is exactly opposite. (Throughout, I gloss case morphemes as 'ALPHA', 'BETA', etc. to facilitate crosslinguistic comparison.) Since post-verbal word order is generally flexible, this means that an identical string of nouns can convey mirror-image events, with event participants swapping roles.

(1) a. P<**um**>atay **ang** balyena **ng** pating.
kill<**AV**> **ALPHA** whale **BETA** shark
'The whale killed the shark.' [Verb_{AV} – $\text{Agent}_{\text{ALPHA}}$ – $\text{Patient}_{\text{BETA}}$]

b. P<**in**>atay **ang** balyena **ng** pating.
kill<**PV**> **ALPHA** whale **BETA** shark
'The shark killed the whale.' [Verb_{PV} – $\text{Patient}_{\text{ALPHA}}$ – $\text{Agent}_{\text{BETA}}$] (Hsieh 2016)

Compare some similar sentences in Georgian (2). Like the other South Caucasian languages (Boeder 2005), Georgian (Shanidze 1953, Aronson 1990) has flexible verb-final word order (Skopeteas et al. 2009) and split-ergative case marking conditioned by tense (Harris 1985, Nash 2017). For instance, if a monotransitive verb is inflected for the future tense, agents are marked by the case suffix *-i* 'ALPHA', and patients by *-s* 'BETA' (2a). In the perfect tense, the morphosyntactic mapping between case and role is reversed (2b).

(2) a. veʃap'**-i** zvigen**-s** mok'lavs.
whale**-ALPHA** shark**-BETA** kill:**FUT**
'The whale will kill the shark.' [$\text{Agent}_{\text{ALPHA}}$ – $\text{Patient}_{\text{BETA}}$ – Verb_{FUT}]

b. veʃap'**-i** zvigen**-s** mouk'lavs.
whale**-ALPHA** shark**-BETA** kill:**PERF**
'The shark has killed the whale.' [$\text{Patient}_{\text{ALPHA}}$ – $\text{Agent}_{\text{BETA}}$ – $\text{Verb}_{\text{PERF}}$]

So, just like Tagalog, the same string of nouns can correspond to mirror-image events, depending on a crucial cue coming from verbal inflection. An important difference between the languages is the dominant word order: Tagalog is verb-initial, while Georgian is verb-final. There are, of course, many other formal syntactic differences between Western Austronesian voice and South Caucasian split ergativity. But at least in a narrow descriptive sense, Georgian can be considered to have *Reverse Philippine* morphosyntactic alignment.

Having introduced this novel typological parallel, the rest of this paper is structured as follows. Section 2 describes the case-marking systems of these languages in a bit more detail. Section 3 introduces a tool from information theory — conditional entropy — that helps make more precise the parallels between Philippine and Reverse Philippine Alignment. Section 4 explains the relevance for these high-entropy case systems for theories of grammatical-role processing, synthesizing previous psycholinguistic findings from Tagalog and Georgian. Section 5 concludes with some directions for future research in comparative Austronesian–Caucasian sentence processing.

2. Grammatical background

The previous examples illustrate a hallmark of Philippine and Reverse Philippine Alignment systems: case is not a very reliable cue to a noun's grammatical role; the case–role mapping is crucially contingent on a dependency with the verb. And indeed for both Western Austronesian and South Caucasian, the full grammatical systems are quite a bit more complex than the first-pass descriptions above.

In Western Austronesian, languages typically have more than two voice categories. A larger paradigm for Tagalog follows, including other grammatical roles and oblique voices. It illustrates how oblique voice categories — including the 'locative' (3c) and 'circumstantial' (3d) voices — signal that an argument other the agent or patient is marked by *ang* 'ALPHA'. Those non-core arguments, outside of their respective voices, are marked with more specialized case particles (*sa* 'GAMMA' for e.g. locatives) or preposition + particle combinations (*para sa* 'DELTA' for benefactees). Note also certain arguments in the same clause might have identical marking: *ng* 'BETA' marks both the agent and patient in oblique voices (3c,d).

(3)

	Verb	Agent	Patient	Locative	Benefactee
a.	Bumili	**ang** bata	**ng** tela	**sa** palengke	**para sa** nanay.
	buy:**AV**	**ALPHA** child	**BETA** cloth	**GAMMA** market	**DELTA** Mother
	'The child bought cloth at the market for Mother.'				
b.	Binili	**ng** bata	**ang** tela	**sa** palengke	**para sa** nanay.
	buy:**PV**	**BETA** child	**ALPHA** cloth	**GAMMA** market	**DELTA** Mother
	'The child bought the cloth at the market for Mother.'				
c.	Binilhan	**ng** bata	**ng** tela	**ang** palengke	**para sa** nanay.
	buy:**LV**	**BETA** child	**BETA** cloth	**ALPHA** market	**DELTA** Mother
	'The child bought the cloth at the market for Mother.'				
d.	Ibinili	**ng** bata	**ng** tela	**sa** palengke	**ang** nanay.
	buy:**CV**	**BETA** child	**BETA** cloth	**GAMMA** market	**ALPHA** Mother
	'The child bought the cloth at the market for Mother.' (Rackowski & Richards 2005)				

These Tagalog data are representative of Philippine Alignment more generally, whereby a very cosmopolitan 'ALPHA' case category is mapped to a specific role indicated by verbal voice morphology. But the great size of the Austronesian family means there is considerable microvariation in the inventory of voice categories and case markers (see Blust 2013). For instance, consider Paiwan (4). In this language, there is dedicated case marker that marks agents outside of the agent voice (*nua* 'DELTA'); there is also a morpheme (*tua* 'BETA') that signals both patients outside of patient voice, and instruments outside of circumstantial voice.

(4)

	Verb	Agent	Patient	Location	Instrument
a.	Qmalup	**a** caucau	**tua** vavuy	**i** gadu	**tua** vuluq.
	hunt:**AV**	**ALPHA** man	**BETA** pig	**GAMMA** mountain	**BETA** spear
	'The man hunts wild pigs in the mountains with a spear.'				

b. Qalupen **nua** caucau **a** vavuy **i** gadu **tua** vuluq.
hunt:**PV** **DELTA** man **ALPHA** pig **GAMMA** mountain **BETA** spear
'The man hunts wild pigs in the mountains with a spear.'

c. Qalupan **nua** caucau **tua** vavuy **a** gadu **tua** vuluq.
hunt:**LV** **DELTA** man **BETA** pig **ALPHA** mountain **BETA** spear
'The man hunts wild pigs in the mountains with a spear.'

d. Siqalup **nua** caucau **tua** vavuy **i** gadu **a** vuluq.
hunt:**CV** **DELTA** man **BETA** pig **GAMMA** mountain **ALPHA** spear
'The man hunts wild pigs in the mountains with a spear.' (Chen & McDonnell 2019)

A hotly debated topic in Austronesian linguistics concerns the formal syntactic properties of the voice system (for an overview, see Chen & McDonnell 2019). An important generalization is that the ALPHA-marked argument is usually definite and highly topical; it also has privileged status in relativization and other filler–gap dependencies. But for present purposes, I abstract away from the interactions between case marking and discourse structure, and remain agnostic as to their relationship with syntactic clause structure.

As for South Caucasian, case marking there is entirely orthogonal to topicality and filler–gap formation. Rather, a clause's case–role mapping is determined by its verb's tense–aspect–mood inflection. TAM categories fall into a few sets, defined in morphological terms (which correspond, at least diachronically, to different syntactic structures; Harris 1985). Examples below use a Georgian ditransitive verb (5) to illustrate how agents, patients, and goals shift case marking across three representative tenses. Note how some case markers (*-i* 'ALPHA' and *-s* 'BETA') can mark a wide variety of roles, whereas others (*-ma* 'GAMMA' and *-is=tʰvis* 'DELTA') are only found on a specific role in a particular tense.

(5)

	Agent	Goal	Patient	Verb
a.	masts'avlebel-**i**	bavʃveb-**s**	ts'ign-**s**	atʃʰvenebs
	teacher-**ALPHA**	children-**BETA**	book-**BETA**	show:**FUT**
	'The teacher will show a book to the children.'			
b.	masts'avlebel-**ma**	bavʃveb-**s**	ts'ign-**i**	atʃʰvena
	teacher-**GAMMA**	children-**BETA**	book-**ALPHA**	show:**AOR**
	'The teacher showed a book to the children.'			
c.	masts'avlebel-**s**	bavʃveb-**is=tʰvis**	ts'ign-**i**	utʃʰvenebia
	teacher-**BETA**	children-**DELTA**	book-**ALPHA**	show:**PERF**
	'The teacher has shown a book to the children.'			

The South Caucasian family is much smaller than Austronesian, exhibiting less microvariation. Georgian's sibling languages mostly behave as in (5) — with some minor differences concerning intransitive subjects, and whether the case pattern in (5b) subsumes that of (5a) (see Harris 1985 on Megrelian and Laz).

3. Conditional entropy and the complexity of case comprehension

While there are significant syntactic differences between Western Austronesian and South Caucasian languages, from the perspective on sentence processing, they have something striking in common. In order to comprehend who did what to whom in a Philippine or Reverse Philippine language, it will very often be necessary to parse a complex grammatical dependency between a noun's case marking and verbal inflection — a dependency totally independent from properties inherent to the verb, like its lexical semantics and argument structure.

One way to make this notion of comprehension-complexity more precise is to apply tools from information theory (cf. Futrell & Hahn 2022). Specifically, *conditional entropy of role given case*, notated H(role|case), is a metric that quantifies in very abstract terms how difficult it is, on average, to process a noun's grammatical role from the evidence of its case morphology. Grammars where each role is uniquely associated with some case morpheme will have very low values for H(role|case); grammars with messier mappings, as in Philippine and Reverse Philippine Alignment, will have higher ones.

The equation for this conditional entropy value is given in (6), which can be unpacked in the following way. Consider every logical combination of roles and cases (agents in case ALPHA, agents in BETA, agents in GAMMA, patients in ALPHA, etc.): $\Sigma_{r\in role,c\in case}$. Calculate the probability of each pairing: P(r,c). Multiply that by the log-transformed conditional probability of that role given that case: log P(r|c). Sum up all such products, and reverse the sign.

(6) $$H(\text{role}|\text{case}) = -\Sigma_{r\in role,c\in case}\, P(r,c) \log P(r|c)$$

An accurate calculation of this hinges on corpus frequencies for every case–role combination. For Georgian at least, we can adopt corpus counts that Foley 2022 inferred from the Georgian National Corpus (Gippert & Tandashvili 2015). (While the GNC is not syntactically parsed, counts of nouns instantiating each grammatical role can be very accurately estimated from the corpus's morphological tags on verbs for tense and argument structure features.) Table 1 gives raw counts for agents, patients, goals, and themes (i.e., unaccusative subjects; these are not described above, but are always marked ALPHA in Georgian).

	Agent	**Patient**	**Theme**	**Goal**
Tense Set 1 (FUT…)	52,709 (ALPHA)	43,143 (BETA)	38,467 (ALPHA)	26,148 (BETA)
Tense Set 2 (AOR…)	68,500 (GAMMA)	75,825 (ALPHA)	48,717 (ALPHA)	36,980 (BETA)
Tense Set 3 (PERF…)	10,181 (BETA)	12,101 (ALPHA)	7,947 (ALPHA)	3,013 (DELTA)

Table 1: Estimated counts of nouns in four grammatical roles across tense sets in Georgian, adapted from Foley 2022. Case marking associated with each role in each tense is given in parentheses.

From these data, it is straightforward to calculate the necessary values of P(r,c) and P(r|c). Thus, I estimate H(role|case) in Georgian to be 1.22 bits, where 1 bit of entropy is the amount of uncertainty associated with predicting an event with two equally likely outcomes. So, on average — abstracting away from all other relevant (and undoubtably important) cues like

discourse context, animacy, and word order — it is easier to predict a fair coin toss (1 bit of entropy) than it is to predict a Georgian noun's grammatical role given its case morphology. As pointed out above (5), certain case morphemes in Georgian are perfectly reliable cues to grammatical role: *-ma* 'GAMMA' always marks agents, and *-is=tʰvis* 'DELTA' always marks goals. Therefore, the bulk of the conditional entropy in this grammatical system comes from the much wider distributions of *-i* 'ALPHA' and *-s* 'BETA'.

What about for Tagalog and Paiwan? Ideally we would consult syntactically parsed corpora for these languages. None are available to me, so instead I tentatively infer values for *P*(r,c) in these languages by manipulating the Georgian corpus data. I did so by mapping all values for Georgian's Tense Set 2 (the most common tenses) to the patient voice (the most common voice in Tagalog, according to Pizarro-Guevara & Garcia 2024), and Set 1 values to the agent voice (the next most common categories). I distributed Set 3 values equally among the locative and circumstantial voices (the least common ones). To account for the greater number of grammatical roles relevant to Western Austronesian, I split counts for Georgian goals equally to derive counts for locatives and circumstances (i.e. benefactees for Tagalog, instruments for Paiwan) in the agent and patient voices. For the oblique voices, I summed up the counts for themes and patients, assuming that every verb in an oblique voice would have a corresponding ALPHA-marked oblique argument; I estimated counts for the other oblique argument (i.e. locatives in the circumstantial voice, and vice versa) to be one quarter the number of agents. Resulting figures are given the following table.

	Agent	**Patient**	**Theme**	**Locative**	**Ben./Instr.**
Agent voice	52,709 (ALPHA/ALPHA)	43,143 (BETA/BETA)	38,467 (ALPHA/ALPHA)	13,074 (GAMMA/GAMMA)	13,074 (DELTA/BETA)
Patient voice	68,500 (BETA/DELTA)	75,825 (ALPHA/ALPHA)	48,717 (ALPHA/ALPHA)	18,490 (GAMMA/GAMMA)	18,490 (DELTA/BETA)
Locative voice	5,090.5 (BETA/DELTA)	6,050.5 (BETA/BETA)	3,973.5 (BETA/DELTA)	10,024 (ALPHA/ALPHA)	1,272.625 (DELTA/BETA)
Circum. voice	5,090.5 (BETA/DELTA)	6,050.5 (BETA/BETA)	3,973.5 (BETA/DELTA)	1,272.625 (GAMMA/GAMMA)	10,024 (ALPHA/ALPHA)

Table 2: Counts of nouns in five grammatical roles across voice categories, estimated for two Western Austronesian languages using the method described in the text. Case values are given for Tagalog (in parentheses on the left; cf. (3)), and Paiwan (on the right; (4)). Case inflection given for themes (i.e. intransitive subjects) is based on descriptions in Chen (to appear).

Caveat lector; the decisions made to derive the frequences above are admittedly ad hoc, and future corpus research or computational analysis is necessary to lend them more credence. Nevertheless, data in Table 2 are sufficient to very roughly estimate *H*(role|case) for these under-resourced Austronesian languages — namely, 1.42 bits for Tagalog and 1.32 bits for Paiwan. So, it seems that these languages' Philippine Alignment systems are, in a technical sense, even more complex to comprehend than Georgian's Reverse Philippine Alignment. For comparison, remapping the data in Table 2 to a canonical Nominative–Accusative Alignment system yields a mere 0.52 bits of conditional entropy; a canonical Ergative–Absolutive system is 0.56 bits.

Note also that *H*(case|role) gives some justification for assigning a common label ('ALPHA') to the Tagalog case marking *ang* (3), the Paiwan marker *a* (4), and the Georgian suffix *-i* (5). For each of these languages, that is the case morpheme with the highest conditional

entropy value — respectively, H(case|ALPHA) comes to 1.02, 1.07, and 0.85 in these languages. Likewise, the morphemes glossed 'BETA' are the second-most entropic morphemes. Morphological case categories across languages are very tricky to compare on formal–representational grounds, given the many ways in which abstract case features are theorized to be assigned syntactically and exponed morphologically. Information theory offers a more objective way to compare language-specific case categories, remaining neutral on the formal mechanisms involved in the computation of their distribution.

4. Incremental processing of high-entropy case systems

The previous section appealed to information theory in exploring a novel conceptual connection between Western Austronesian's Philippine Alignment system and South Caucasian's Reverse Philippine Alignment system. Both types of grammars result in very high conditional entropy values for grammatical role given case morphology.

Let us now consider how transitive clauses are processed word-by-word in these high-entropy languages. I take inspiration from Wagers et al.'s (2018) dissection of the unfolding cues in Chamorro relative clauses, and Bornkessel-Schlesewsky & Schlesewsky's (2009b) extended Argument Dependency Model (eADM) for the neurolinguistic processing of grammatical roles. In comparing Philippine and Reverse Philippine languages, key will the position of the verb. For any given clause, this word will bear the crucial cue (viz. voice in Tagalog or Paiwan; tense in Georgian) that disambiguates what role is signaled by the highest-entropy cases.

Philippine-type languages are verb initial. So, just one word into a sentence — e.g., at the position marked ① in the following Tagalog example (7) — the comprehender is faced with several significant processing tasks. They must access the lexical entry of the verb root, which conveys the event-concept which the sentence expresses: here, a 'killing' event. That in turn might evoke expectations about likely event participants; given entailments of this verb, both the agent and patient are very likely to be animate entities (cf. Sauppe's 2016 eye-tracking evidence in Tagalog). The comprehender must also process the verb's argument structure: here, monotransitive, which licenses them to project syntactic positions to accommodate the upcoming agent and patient. And finally, they must process the voice morphology: here, the agent-voice infix, which through a bit of grammatical calculation leads them to predict with certainty that the agent will be marked ALPHA and the patient BETA.

(7) Pumatay ① ang balyena ② ng pating. ③
kill:AV ALPHA whale BETA shark
'The whale killed the shark.' (Hsieh 2016)

The next word the comprehender will encounter is a noun, at position ②. They must process the noun's lexical semantics, and its case morphology. According to eADM, both lexical semantic and morphosyntax cues to grammatical role are processed in parallel. Since *whale* refers to an animate entity, that makes it a canonical agent: merely on semantic grounds, then, there is good evidence that this noun occupies the transitive subject position that had been previously projected. This noun is also first in the sentence; linear prominence, qua gradient cue for evaluating role prototypicality, is also a heuristic predicted by eADM to influence processing. At the same time, a stream of morphosyntactic processing also arrives at the conclusion that this noun is the agent. The comprehender retrieves the voice-feature of the previously encountered

verb: agent voice entails that ALPHA marks the agent role. The noun's role successfully processed, the comprehender can integrate the meaning of the incomplete sentence: it refers to a killing event caused by a whale.

Finally, at position ③, the comprehender encounters the second noun. The same tasks unfold, likely aided by the fact that most of the sentence has already been processed. Having semantically and syntactically integrated all three words of the sentence, the comprehender now knows who did what to whom: the whale killed the shark.

Similar incremental processing routines will unfold for every sentence in Tagalog (and other Philippine-Alignment languages). But how useful is this type of theorization for understanding empirical sentence-processing behavior in this type of language? To date, a small but rapidly growing number of psycholinguistic studies have been run on Philippine-type languages, and already a few generalizations seem clear (see Pizarro-Guevara & Garcia 2024 for a review). Two sentence-processing constraints are described as interacting in Tagalog: one enforces a preference for Agent-First word orders, and another for ALPHA-Last orders. Hsieh's (2016) naturalness-judgement results illustrate quite clearly (Table 3). This study found that the best-rated sentences were those which satisfy both constraints: namely, ones with Verb_{PV} – $\text{Agent}_{\text{BETA}}$ – $\text{Patient}_{\text{ALPHA}}$ word order. The least natural sentences were those that violated both constraints: Verb_{PV} – $\text{Patient}_{\text{ALPHA}}$ – $\text{Agent}_{\text{BETA}}$.

Tagalog sentence frame	**Naturalness rating (out of 7)**
Verb_{AV} – $\text{Agent}_{\text{ALPHA}}$ – $\text{Patient}_{\text{BETA}}$	good (≈ 4.7)
Verb_{AV} – $\text{Patient}_{\text{BETA}}$ – $\text{Agent}_{\text{ALPHA}}$	good (≈ 4.7)
Verb_{PV} – $\text{Agent}_{\text{BETA}}$ – $\text{Patient}_{\text{ALPHA}}$	very good (≈ 5.4)
Verb_{PV} – $\text{Patient}_{\text{ALPHA}}$ – $\text{Agent}_{\text{BETA}}$	bad (≈ 4.2)

Table 3: Summary of results from Hsieh's (2016) acceptability study on transitive sentences in Tagalog, manipulating word order and voice.

The incremental story sketched above helps us understand the Agent-First constraint. The first noun to appear in a sentence occupies an intrinsically prominent linear position, according to the prominence scales by which eADM evaluates arguments' role prototypicality. Indeed, Agent-First is believed to be a ubiquitous processing constraint, if not a universal one, perhaps rooted deeply in the way that human cognition perceives events (Bornkessel-Schlesewsky & Schlesewsky 2014, Sauppe et al. 2023).

But why might there be an ALPHA-Last constraint in Tagalog? I offer three suggestions. First is a formal syntactic explanation. The ALPHA-marked argument might originate in a particular right-peripheral position in the clause, or else move to that position in order to be assigned case (cf. Kroeger 1991, Guilfoyle et al. 1992). An extra movement operation would thus be necessary to derive BETA-last word orders, which in turn might result in a cost to processing or acceptability.

Second is an explanation related to the Entropy Reduction Hypothesis (Hale 2003, 2006). This states that, in general, the processing cost of a word is proportional to the degree to which it reduces the uncertainty about upcoming words and syntactic structures. As noted above, Tagalog's ALPHA case category has the language's highest H(case|role) value. Perhaps clauses where nouns are ordering according to increasing case–role entropy (i.e., where BETA comes

before ALPHA) are easiest to process. The intuition here is that ALPHA-Last order provides a 'smoother landing' for the comprehender, dealing with the most complex case category only once every other box has been checked, and the majority of the event has been processed. Formalizing this intuition explicitly in terms of Entropy Reduction will require future research, perhaps involving a probabilistic context-free grammar for Tagalog built from sufficiently rich corpus data.

The third possibility is that ALPHA-Last is ultimately a consequence of eADM's mechanism for assessing role prototypicality. Just as the theory posits prominence hierarchies for animacy and word order, it also posits one for case morphology: with nominative outranking accusative, and ergative outranking nominative/absolutive (Bornkessel-Schlesewsky & Schlesewsky 2009b:329). While theory does not explicitly state how to make this case hierarchy general enough to accommodate alignment systems other than canonical Nominative–Accusative and Ergative–Absolutive ones, I suggest that it can be derived from the probability that an agent will be inflected for a particular case category: that is, from *P*(case|agent). For Tagalog, the frequencies inferred above (Table 2) entail *P*(BETA|agent) to be 0.60, and *P*(ALPHA|agent) to be 0.40. Thus BETA outranks ALPHA in prominence for this language. So, all else equal, eADM predicts that an ALPHA-marked noun directly following the verb to be a slightly less attractive agent than a BETA-marked noun in this position; hence, ALPHA-Last.

Now, let us turn to the incremental processing of a Reverse Philippine sentence. The key difference is that all of the information the comprehender can glean from the verb — its lexical semantics, its argument structure, and its tense morphology that crucially disambiguates which roles the certain cases signal — is withheld until the end of the clause. Consider the following Georgian sentence.

(8) veʃap'-i ① zvigen-s ② mok'lavs. ③
whale-ALPHA shark-BETA kill:FUT
'The whale will kill the shark.'

Upon encountering the first word, at position ①, all the comprehender knows is that the sentence has at least one core argument. ALPHA is the language's most entropic case category, so there is quite a bit of uncertainty about what could come next. Insofar as any syntactic predictions are licensed about the rest of the sentence, they may be due to the noun's lexical semantics: *whale* is high-animacy, so a canonical agent; a transitive clause structure is therefore a reasonable prediction. The second word, a BETA-marked noun ②, is compatible with this prediction, but it does not rule out alternatives. That is because this noun in fact could turn out to be the agent, and the previous noun to be the patient; compare the perfect-tense (2b). (A third grammatical possibility is that these two nouns are the objects of a ditransitive verb: imagine a version of (5b) where the GAMMA-marked agent is either postverbal or *pro*-dropped.) Finally, at the verb ③, the comprehender encounters everything necessary to process who did what to whom. All of their predictions about the clause's argument structure and case–role mapping will be either confirmed or foiled. The lexical semantics of all three words can be integrated to arrive at an event of shark-killing caused by the whale.

Empirical work in Georgian psycholinguistics is also nascent. But relevant here are reading-time data reported by Foley (2024). Tracking the incremental comprehension of sentences like (8), a key finding is that verbs in $\text{Patient}_{\text{ALPHA}}$ – $\text{Agent}_{\text{BETA}}$ – $\text{Verb}_{\text{PERF}}$ clauses are

much harder to process than verbs in minimally different contexts (Table 4). (Similar results obtain in the RTs for Skopeteas et al.'s 2012 timed acceptability-judgement experiment.)

Georgian sentence frame	**Processing time of verb (RT)**
$\text{Agent}_{\text{ALPHA}}$ – $\text{Patient}_{\text{BETA}}$ – Verb_{FUT}	fast (7.16 log ms)
$\text{Patient}_{\text{BETA}}$ – $\text{Agent}_{\text{ALPHA}}$ – Verb_{FUT}	fast (7.18 log ms)
$\text{Agent}_{\text{BETA}}$ – $\text{Patient}_{\text{ALPHA}}$ – $\text{Verb}_{\text{PERF}}$	fast (7.16 log ms)
$\text{Patient}_{\text{ALPHA}}$ – $\text{Agent}_{\text{BETA}}$ – $\text{Verb}_{\text{PERF}}$	slow (7.26 log ms)

Table 4: Summary of results from Foley's (2024) self-paced reading study in Georgian, manipulating word order and tense of verb-final monotransitive clauses

This is eerily similar to the Tagalog findings, whereby a $\text{Patient}_{\text{ALPHA}}$ – $\text{Agent}_{\text{BETA}}$ string likewise impedes processing, even though the verb appears first in that language. And the hypotheses laid out above to explain Tagalog's ALPHA-Last constraint can each be plausibly extended to Georgian. First, the offensive $\text{Patient}_{\text{ALPHA}}$ – $\text{Agent}_{\text{BETA}}$ – $\text{Verb}_{\text{PERF}}$ word order arguably must be derived with an extra step of movement targeting the patient (cf. Lomashvili & Harley 2011), so a formal syntactic explanation is plausible. Second, H(case|ALPHA) in Georgian is higher than H(case|BETA), so the intuition that nouns should be ordered in increasing comprehension-entropy applies here. And third, the corpus data in Table 1 show that P(ALPHA|agent) > P(BETA|agent) for Georgian, so an extension of eADM's case-prominence hierarchy is also a possible account of this processing asymmetry.

5. Conclusion

In sum, despite many typological differences between the Philippine Alignment pattern found in Western Austronesian languages and the Reverse Philippine pattern in South Caucasian, these languages are uncanny foils for each other when it comes to sentence processing. They differ most notably in the default position of the verb — at the beginning or end of the clause — which has profound consequences for the incremental identification of who did what to whom. Besides identifying the event concept and constraining the clause's syntactic argument structure, the verb in Philippine and Reverse Philippine languages also bears a cue that crucially unlocks the grammatical/thematic role of most nouns. Even given the limited amount of sentence-processing data currently available for Tagalog and Georgian, a striking parallel is apparent. In both languages, it appears that the hardest types of simple transitive clauses to process are those where nouns are ordered $\text{Patient}_{\text{ALPHA}}$ – $\text{Agent}_{\text{BETA}}$, where H(role|ALPHA) > H(role|BETA). This alone warrants targeted comparison between the languages, as a promising avenue to advance typologically general theories of sentence processing.

One place to develop this line of research, perhaps untangling the hypotheses laid out above, is in sentences where case–role mappings are globally ambiguous. In Tagalog monotransitives, both the agent and patient can be marked BETA in the oblique voices (9a,b) and also in the recent perfective construction (9c). In Georgian ditransitives, both patients and goals will be marked BETA in the future tense (9d). Carefully manipulating word order, animacy, and information structure in sentences like these might shed light on the precise ways in which linguistic cues guide incremental comprehension of grammatical roles.

(9) a. Binilhan **ng** bata **ng** tela ang palengke para sa nanay.
buy:**LV** **BETA** child **BETA** cloth ALPHA market DELTA Mother
'The child bought the cloth at the market for Mother.' (Rackowski & Richards 2005)

b. Ibinili **ng** bata **ng** tela sa palengke ang nanay.
buy:**CV** **BETA** child **BETA** cloth GAMMA market ALPHA Mother
'The child bought the cloth at the market for Mother.' (Rackowski & Richards 2005)

c. Kapangunguha pa lamang **ng** bata **ng** mga mangga
gather:**REC.PERF** yet only **BETA** child **BETA** PL mango
'The child has just gathered some/the mangoes.' (Kroeger 1991)

d. masts'avlebel-i bavʃveb-**s** ts'ign-**s** atʃʰvenebs
teacher-ALPHA children-**BETA** book-**BETA** show:**FUT**
'The teacher will show a book to the children.'

Comparative sentence processing is a growing branch of psycholinguistics (Sauppe et al. 2023, Polinsky 2023), with much potential to illuminate the connections between grammar, typology, and language-comprehension mechanisms. Experimental techniques suitable for under-resourced are rapidly improving and becoming more accessible (Wagers & Chung 2023). Austronesian languages — and not just ones with Philippine Alignment — offer many opportunities for fruitful psycholinguistic research, through language-internal investigation and in comparison within the family and beyond.

References

Aronson, Howard I. 1990. *Georgian: A reading grammar*. Bloomington: Slavica Publishers.

Blust, Robert. 2013. *The Austronesian Languages*. Canberra: Asia-Pacific Linguistics.

Bornkessel-Schlesewsky, Ina, and Matthias Schlesewsky. 2009a. *Processing syntax and morphology: A neurocognitive perspective*. Oxford: Oxford University Press.

Bornkessel-Schlesewsky, Ina, and Matthias Schlesewsky. 2009b. The role of prominence information in the real-time comprehension of transitive constructions: A cross-linguistic approach. *Language and Linguistics Compass* 3: 19-58.

Bornkessel-Schlesewsky, Ina, and Matthias Schlesewsky. 2014. Scales in real-time language comprehension: A review. *Scales and hierarchies: A cross-disciplinary perspective*, eds. Ina Bornkessel-Schlesewsky, Andrej L. Malchukov, and Marc Richards. Berlin: De Gruyter Mouton.

Chen, Victoria. To appear. Philippine-type "voice" affixes as A'-agreement markers: Evidence from causatives and ditransitives. *Natural Language and Linguistic Theory*.

Chen, Victoria, and Bradley McDonnell. 2019. Western Austronesian voice. *Annual Review of Linguistics* 5:173-195.

Ferreira, Fernanda, and Zhuang Qiu. 2021. Predicting syntactic structure. *Brain Research* 1770:147632.

Foley, Steven. 2022. Quantifying performance constraints on case-alignment typology with information theory. Poster presented at 35th Annual Conference on Human Sentence Processing. Santa Cruz, California.

Foley, Steven. 2024. Case, word order, and the incremental disambiguation of grammatical roles in Georgian. Ms., University of Georgia, Athens, Ga.

Futrell, Richard, and Michael Hahn. 2022. Information theory as a bridge between language function and language form. *Frontiers in Communication* 7:657725.

Gippert, Jost, and Manana Tandashvili. 2015. Structuring a diachronic corpus: The Georgian National Corpus project. *Historical Corpora: Challenges and perspectives*, 305-322. Tübingen: Narr Verlag.

Guilfoyle, Eithne, Henrietta Hung, and Lisa Travis. 1992. Spec of IP and spec of VP: Two subjects in Austronesian languages. *Natural Language and Linguistic Theory* 10:375-414.

Hale, John. 2003. The information conveyed by words in sentences. *Journal of Psycholinguistic Research*, 32:101-123.

Hale, John. 2006. Uncertainty about the rest of the sentence. *Cognitive Science*, 30:643-672.

Harris, Alice C. 1985. *Diachronic syntax: The Kartvelian case.* Orlando: Academic Press.

Hsieh, Henrison. 2016. Prosodic indicators of phrase structure in Tagalog transitive sentences. *Proceedings of the Austronesian Formal Linguistics Society* 23, 111-122. Asia-Pacific Linguistics, Canberra, Australia.

Kroeger, Paul. 1991. Phrase structure and grammatical relations in Tagalog. Doctoral Dissertation, Stanford University.

MacWhinney, Brian, and Elizabeth Bates. 1989. *The Crosslinguistic Study of Sentence Processing*. Cambridge: Cambridge University Press.

Lomashvili, Leila, and Heidi Harley. 2011. Phases and templates in Georgian agreement. *Studia Linguistica* 65:233-267.

Nash, Léa. 2017. The structural source of split ergativity and ergative case in Georgian. *The Oxford Handbook of Ergativity*, eds. Jessica Coon, Diane Masssam, and Lisa Demena Travis, 175-200. Oxford: Oxford University Press.

Pizarro-Guevara, Jed Sam, and Rowena Garcia. 2024. Philippine psycholinguistics. *Annual Review of Linguistics* 10:145-167.

Polinsky, Maria. 2023. Experimental syntax and linguistic fieldwork. *The Oxford Handbook of Experimental Syntax*, ed. John Sprouse, 97-126. Oxford: Oxford University Press.

Rackowski, Andrea, and Norvin Richards. 2005. Phase edge and extraction: A Tagalog case study. *Linguistic Inquiry* 36:565-599.

Sauppe, Sebastian. 2016. Verbal semantics drives early anticipatory eye movements during the comprehension of verb-initial sentences. *Frontiers in Psychology* 7:95.

Sauppe, Sebastian, Åshild Næss, Giovanni Roversi, Martin Meyer, InaBornkessel-Schlesewsky, and Balthasar Bickel. 2023. An agent-first preference in a patient-first language during sentence comprehension. *Cognitive Science* 47:e13340.

Sauppe, Sebastian, Caroline Andrews, and Elisabeth Norcliffe. 2023. Experimental research in cross-linguistic psycholinguistics. *Routledge Handbook of Experimental Linguistics*, eds. Sandrine Zufferey and Pascal Gygax, 156-172. New York: Routledge.

Shanidze, Akaki. 1953. *Kartuli enis gramat'ik'is sapudzvlebi* [Fundamentals of Georgian grammar]. Tbilisi: Tbilisi State University Press.

Skopeteas, Stavros, Gisbert Fanselow, and Rusudan Asatiani. 2012. Case inversion in Georgian: Syntactic properties and sentence processing. *Case, Word Order and Prominence: Interacting cues in language production and comprehension*, eds. Monique Lamers and Peter de Swart. Dordrecht: Springer.

Skopeteas, Stavros, Caroline Féry, and Rusudan Asatiani. 2009. Word order and intonation in Georgian. *Lingua* 119:102-127.

Wagers, Matt, Manuel F. Borja, and Sandra Chung. 2018. Grammatical licensing and relative clause parsing in a flexible word-order language. *Cognition* 178:207-221.

Wagers, Matt, and Sandra Chung. 2023. Language processing experiments in the field. *The Oxford Handbook of Experimental Syntax*, ed. John Sprouse, 491-512. Oxford: Oxford University Press.