Transitivity in Toba Batak and Tagalog
Author(s): Fay Wouk
Proceedings of the Tenth Annual Meeting of the Berkeley Linguistics Society (1984), pp. 274-284

Transitivity in Toba Batak and Tagalog

Fay Wouk, UCLA

Western Austronesian languages typically have systems for marking verbs in which the verbal morphology gives information about the case role of one particular argument-NP. That NP is often singled out by position or by a prenominal particle, and it has certain syntactic peculiarities as well. The whole system is usually called a focus system, and the NP whose role is marked on the verb is most often called subject, topic or focused NP. With one-argument verbs the "topic" is the only available NP, either agent or experiencer. When the verb has two or more arguments, a choice must be made between agent/experiencer, patient, and possibly one or more obliques, such as recipient, beneficiary, instrument, or location. The question of which NP is chosen for this distinction has never been adequately answered, although three possible explanations exist. In this paper I will briefly consider voice and focus as explanations, and show that they are incorrect. I will present discourse data that supports a modified version of the third explanation, the transitivity hypothesis.

The two languages I studied are Toba Batak of Sumatra, Indonesia, and Tagalog of the Philippines. One noun phrase is marked as "topic" or "focused" in each clause. In Batak this NP is marked by position, immediately following the VP, while in Tagalog it is preceded by the article ang. Verbal morphology gives information about the role of this "focused" NP. In Tagalog the morphemes um-, mag-, mang- indicate that the "topic" is the actor, i-, ipag-, ipang- indicate that the "topic" is patient, benefactive, or instrumental, -an, -in indicate that the "topic" is patient or direction, and pag-an, ka-an indicate that the "topic" is locative. In Batak, a prefix indicates either the "topic" is the actor (mang-) or the "topic" is a non-actor (i-), while in the non-actor form, the suffix -hon indicates that the "topic" is not a patient, but rather some sort of oblique, often recipient or beneficiary.

The use of the term "subject" for the NP whose role is registered on the verb suggests a choice of active vs passive that is often explained in terms of thematicity. Thompson (forthcoming) suggests that the use of English passive "seems to be based on the discourse structuring principles of thematic and inter-clausal continuity" (p.18). She concludes that

passive subjects tend to be thematic, while passive agents are non-thematic, and that the function of passive is to keep thematic elements in the subject slot. "Topic" also suggests thematicity, while "focus" most often refers to new information. There are problems with these analyses. The NP so pointed out is neither "focused" in the sense of presenting new information nor is it the "topic" in the sense of theme. The most common so-called "topics" are patients. It has been noted (Givon, 1979) that patients are statistically less likely to be human and definite than non-human and indefinite, while the opposite is true of agents. However, humans and definites are more likely to be thematic than non-humans. Thus, agents are more likely to be thematic than patients, and a theme oriented system will generally have more agents then patients as subject. Furthermore, it has long been noted that at least in Tagalog it is definite patients which become "topics", so clearly new information is not being presented. In terms of discourse function, the system in question is neither a focus system nor a voice system. I will use the more neutral term trigger in place of topic or focus throughout this paper. The overall system will be called a trigger system, and the NP will be called the trigger, since its case role triggers the verbal morphology. Clauses with actors for triggers will be referred to as +AT, while those with patients or other participants as triggers will be referred to as -AT.

Recently, Hopper & Thompson (1980) have suggested a new explanation of trigger choice on the basis of their scalar transitivity hypothesis. According to this hypothesis, the level of transitivity of a given clause is not just a result of the number of arguments which the verb has, but rather of a combination of factors, or parameters, "all concerned with the effectiveness with which an action takes place," such that verbs with the same number of arguments could be higher or lower in transitivity with respect to each other. Hopper & Thompson found that these parameters tended in fact to co-vary, so that a clause that was high in one would be high in several others. A given clause will be more transitive if it ranks higher on a number of these parameters, and less transitive if it ranks lower. The parameters that they consider relevant to transitivity are given in the following table.

Table 1	high	low
participants	two or more	one
kinesis	action	non-action
aspect	telic	atelic
punctuality	punctual	durative
volitionality	volitional	non-volitional
affirmation	affirmative	negative
mode	realis	irrealis
agency	high in potency	low in potency
affectedness of P	totally affected	unaffected
individuation of P	highly indiv.	non-indiv.

Highly individuated patients are opposed to less individuated patients in the following way:

Table 2 individuated	non-individuated
proper	common
human, animate	inanimate
concrete	abstract
singular	plural
count	mass
referential, definite	non-referential

Hopper & Thompson also show that cross-linguistically, morphosyntax is often sensitive to the overall level of transitivity of a clause, rather than just to the number of arguments present. Morphological marking in many languages differentiates "transitive" and "intransitive" on bases other than number of arguments, singling out e.g. past or perfect two-argument verbs only, or two-argument verbs with individuated patients only, rather than marking all two argument verbs as "transitive". Their study includes a number of Austronesian languages in which the referential status of the patient is more important than the presence or absence of a patient with regards to morphological differentiation from "intransitive" or one-argument verbs.

Hopper & Thompson go on to argue that in discourse, high transitivity correlates with foreground, and low transitivity with background. That is, they claim that in the main event line of a narrative text, there will be a preponderance of high transitivity clauses, while in the background and descriptive material there will be a preponderance of low transitivity clauses. They state that in Tagalog +AT clauses (with no patient or an indefinite patient and imperfective aspect) correlate with background, and -AT clauses (with a definite patient and

perfective aspect) correlate with foreground. Hopper (1979a, 1979b) also applies this analysis to Old Javanese and Malay, and claims that foreground, perfective aspect and definite patients all correlate with the -AT verb forms.

I examined their hypothesis with respect to both Tagalog and Toba Batak. These two languages are generally believed to belong to the same first order subgroup of Austronesian, although they almost certainly do not belong to the same immediate subgroup. Thus similarities between them will be of interest for both synchrony and diachrony.

The Toba Batak texts (4 long and 11 short) were all oral texts, collected during a UCLA field methods course. They were produced by a single speaker, and represent a variety of text types, including narrative, procedural, descriptive, and expository. The Tagalog texts (8 in number) included both written (formal and informal), oral and dictated texts. They were from three different speakers, and all are basically narrative in nature.

For every potentially two-argument verb in the corpus I tabulated the results for foreground vs background, aspect, and syntactic and semantic individuation of the patient. In determining foreground vs background, I applied Hopper & Thompson's definitions of foreground and background as much as possible. In the non-narrative texts there was no sequential event line, so I divided clauses into event vs non-event. To investigate aspect, I first coded all irrealis clauses in both languages, and non-finite ones in Tagalog in such a way as to exclude them from consideration, since the aspectual distinction is neutralized in those cases. The remaining Tagalog clauses all had either perfective or imperfective morphology. In Batak I originally distinguished four categories, based on a combination of semantics and morphology: general past, perfective, habitual and progressive. The general past and progressive categories proved too small to use, so I combined habitual and progressive as imperfective, and general past and perfective as perfective. Syntactic individuation of the patient was determined by coding, no patient or a bare noun being contrasted with an individuated NP, one that was either modified in some way or was anaphoric to a previously modified NP. Semantic individuation was based on the referential status of the patient (definite, specific, non-referential, null) as determined by first or prior mention. I suspected that the presence of a verb of

cognition or speech was relevant, as they appeared on preliminary examination to be preponderantly -AT, so I tabulated that information as well. There are a number of situations in which trigger choice is not free. In both Batak and Tagalog within a relative clause the NP coreferential with the head must be the trigger. In Batak, typically a predicate initial language, topicalized initial NPs are found, but only the trigger NP can be found in this position. These clauses I excluded. After tabulating all these factors, wherever data sets were sufficiently large I analyzed the results using BMDP log-linear analysis, a program that determines the presence and statistical significance of interaction between factors in accounting for variation. Table 3 gives the total number of clauses for each language, and shows the high frequency of -AT constructions.

Table 3 Clause Totals

	Total	+AT	-AT
Tagalog	222	60	162
Batak	218	71	147

Table 4 gives the figures for verbs of cognition and speech.

Table 4 Verbs of Cognition and Speech

		C&S verbs	not C&S
Tagalog	+AT	6	54
	-AT	40	122
Batak	+AT	2	69
	-AT	50	97

As Table 4 shows, the vast majority of verbs of cognition and speech had -AT morphology. This proved to be statistically significant in both languages, with a probability of chance occurrence of .0 in Batak, and .0387 in Tagalog. All later tests were done both considering all clauses and excluding the clauses with verbs of cognition and speech. Since the presence of these verbs did not correlate significantly with any of the other relevant factors for trigger choice, I concluded that this should be considered a separate case of almost completely grammaticized trigger choice. In this paper, therefore, I will not discuss the further results that included verbs of cognition and speech.

Foreground and background proved to have absolutely no correlation with trigger choice, either

statistically significant or otherwise, as the figures in Table 5 show.

Table 5 Trigger Choice & Foreground/Background

		Foreground	Background
Tagalog	+AT	34	20
	-AT	63	59
Batak	+AT	28	41
	-AT	50	47

In both languages -AT is almost evenly split between foreground and background. +AT is less evenly split, but the difference is still fairly small, and the greater quantity is background in Batak but foreground in Tagalog. Foreground and background are terms most appropriate to narrative, so it is not surprising that there is no correlation in Batak, where the texts were of mixed genre, but it is somewhat surprising that there is no correlation in Tagalog, where all the texts were narratives. If -AT is indeed the high transitivity form, then Hopper & Thompson's hypothesis that high transitivity correlates with foreground does not hold up. Their hypothesis assumes that the basic type of discourse is narrative. It is more likely that conversation is the basic type of discourse, and conversation presumably does not have the same kind of foreground vs background distinction that narrative does. If transitivity marking has a discourse function, that function should be relevant to the needs of conversation, not of narrative.

Aspect did prove to be statistically significant in Batak, but not in Tagalog. The figures are given in Table 6. For interpretation, the irrealis and non-finite columns may be ignored.

Table 6 Trigger Choice and Aspect

		irrealis	imperfect	perfect	non-finite
Tagalog	+AT	4	9	31	10
	-AT	17	22	73	10
		irrealis	imperfect	perfect	
Batak	+AT	21	30	18	
	-AT	27	21	49	

In Tagalog the ratio of imperfect to perfect is the same in +AT and -AT, approximately 1-to-3 and it is not clear from this data that aspect contributes to trigger choice. In Batak aspect proved significant with a probability of .0088. The probability measurement means there is only .0088 possibility of

chance occurrence of a correlation between trigger choice and aspect, but does not indicate in what way trigger choice correlates with aspect. An examination of the raw figures shows that +AT is approximately 65% imperfective, and -AT is approximately 70% perfective. However, there is also a statistically significant (.0403 probability of chance occurrence) correlation between trigger choice and the interaction of aspect and patient syntactic status. This means that part of the apparent effect of aspect on trigger choice is likely to be due to the way aspect correlates with patient syntactic status.

I will argue that -AT correlates highly with individuated patients, that is patients whose presence and identity are of sufficient importance to the situation described for them to be modified or identified in some way, while +AT correlates with non-individuated patients. It is generally the case that in perfective actions the patient tends to be individuated, whereas with habitual actions the patient is often not individuated and indeed non-referential. It seems possible to me that the observed correlation between aspect and trigger choice might be the result of the small number of perfective actions in which the patient is not individuated, and the large number of habitual actions in which the patient is not individuated. In fact, In Batak 62% of all +AT realis verbs are habitual, as opposed to only 23% of all -AT realis verbs. Indeed, the transitivity hypothesis predicts that aspect and patient status are not independent, and the data from Batak bears this out. Thus, there is no evidence that aspect contributes directly to trigger choice in either language.

Interestingly, there is a statistically significant correlation (probability .0) between aspect and foreground/background in Tagalog, as predicted and claimed by Hopper, although not in Batak, as Table 7 shows.

Table 7 Foreground/Background & Aspect

		irrealis	imperfect	perfective	non-finite
Tagalog	F	2	8	86	1
	B	19	23	18	19
		irrealis	imperfect	perfective	
Batak	F	10	25	40	
	B	38	26	27	

In Tagalog, foreground verbs are predominantly perfective (91.5%), while background verbs are more or

less evenly divided among aspects. However, in Batak perfective is only slightly more common than imperfective in foreground (61.5%). It might be argued that this reflects the fact that the Batak texts are not all narratives. However, a separate count of only narrative texts did not show any better correlation. This difference between Batak and Tagalog is interesting in view of the fact that aspect is obligatorily marked in Tagalog, but optionally marked in Batak. Perhaps a strong correlation between foreground and perfective aspect can only be expected when aspectual marking is obligatory.

Patient individuation proved to be the most important factor in both Batak and Tagalog trigger choice. For both languages, both syntactic and semantic criteria proved statistically significant. The two languages proved to differ with respect to whether correlation was better with semantic or syntactic criteria. Table 8 gives results based on syntactic criteria, and Table 9 gives results based on semantic criteria.

Table 8 Patient Syntactic Status

		no patient	unind.patient	indiv.patient
Tagalog	+AT	15	17	22
	-AT	0	13	109
Batak	+AT	24	33	12
	-AT	6	3	88

Table 9 Patient Semantic Status

		no ref	non-ref	spec	def
Tagalog	+AT	15	20	14	5
	-AT	0	16	11	95
Batak	+AT	24	38	5	2
	-AT	6	41	17	33

For Tagalog the syntactic criterion (individuated patient) seems less useful. Close to one sixth of the individuated patients are in +AT clauses, and unindividuated patients are almost evenly split between the two clause types. The semantic criterion gives much better results. 95% of all definite patients are found in -AT clauses. Specific and non-referential patients are problematic, since they are split between clause types. One possible explanation would be in terms of aspect. That is, we might expect that when a patient is not definite, perfective versus non-perfective aspect will be the

deciding criterion. This did not prove to be the case. Indefinite patients are fairly randomly distributed with respect to aspect.

For Batak the syntactic criterion seems a better predictor. There are, after all, more non-referential patients in -AT clauses than there are definite ones. There are, however, very few -AT clauses with no patient or an unindividuated patient, and only 12 +AT clauses with an individuated patient. Again, aspectual differences cannot explain the cases which do not go according to prediction. In both languages there is a certain amount of variation that cannot be explained with reference to patient status.

There is a simple explanation for the variation, in both Batak and Tagalog. I propose that -AT morphology does not directly correlate with any observable phenomenon. It correlates directly with the speaker's choice, or judgement, that a certain referent expressed by a patient NP is relevant to the discourse. It is something that a listener should pay some attention to in evaluating a speaker's utterance. I propose to refer to these NP with the term salient. A salient patient is simply one whose referent the speaker judges to be of some importance, and this fact is signalled by using a -AT structure. Thus, all thematic referents (highly continuous) will be salient, but the reverse is not true. A referent can be salient, but only present in a short portion of the text, or indeed only mentioned once.

In both languages, salience correlates highly with definiteness, zero anaphora and syntactic elaboration, which are other means of signalling the relative importance of an NP. +AT morphology correlates most highly with absence of a patient, and next with minimal individuation. The correlation is not 100%, however, because the process is not a grammatical one. There can be variation even at the two ends of the scale, definiteness and absence of patient. But most of the variation comes in the middle ground, among the non-definite patients. This is true in both Batak and Tagalog.

The greater predictive value of definiteness for Tagalog and syntactic elaboration for Batak probably has more to do with the difference in text types than a difference between languages. Participants in a narrative will generally be referential if they have any importance. This is not necessarily true of non-narratives.

I have attempted to show that -AT structures correlate with salient patients. The use of

descriptive material and zero anaphora represent speaker judgements as to the greater salience of those participants, in comparison with participants that are neither of sufficient interest to describe, nor sufficiently in the consciousness of the hearer to express with zero anaphora. I will now try to explain why that should be the case. I claim that the grammatical function of the trigger system is to distinguish intransitive (or low in transitivity) from transitive (or high in transitivity). This system is then used in discourse to signal the salience of the patient, since high transitivity generally correlates with salience of the patient, and low transitivity correlates with no patient or non-salience of the patient.

There are two kinds of evidence that the grammatical function of the trigger system is to indicate level of transitivity in Hopper & Thompson's sense. The first kind has to do with Hopper & Thompson's transitivity parameters. Several of them co-vary with -AT. -AT verbs usually have two or more participants, have a patient which is syntactically or semantically individuated, have perfective aspect, and involve volition. They are more often realis, affirmative and punctual than irrealis, negative and durative. This is not to claim that intransitive verbs, or +AT verbs cannot have these characteristics. In fact, it is even possible that lower transitivity verbs will have these characteristics more often than they will not. But the opposite characteristics - less than one participant, an unindividuated patient, imperfective aspect, etc. - will be true of a larger proportion of all +AT verbs, and of a smaller proportion of all -AT verbs. And this is indeed the case in my data. The second kind of evidence has to do with the actual morphology. The morphology used for +AT in Tagalog overlaps with the morphology of one-argument verbs. One-argument verbs usually take the -um- infix, and a few take mag- or mang- prefixes. While this is no longer true in Batak, there are a few one-argument verbs which show an alternation between initial /m/ in the indicative and /p/ in the imperative and causative. The /m/ is probably a historical relic of -um-. It is also true that other languages in the same first order subgroup as Batak and Tagalog, but which do not belong to the same immediate subgroup as Tagalog, such as Old Javanese and Chamorro (Fox, 1982), use similar morphology for +AT two-argument verbs and for at least some one-argument verbs. It seems very likely that at

an earlier stage Batak did so as well.

As I mentioned before, Toba Batak and Tagalog almost certainly belong to the same first order subgroup of Austronesian, Malayo-Polynesian. They also probably belong to the same second order sub-group, Western Malayo-Polynesian, but to different branches of that sub-group. On the basis of this study, I tentatively reconstruct a syntactic stage (corresponding to Proto-Western Malayo-Polynesian) in which -AT morphology was used for clauses that were high in transitivity, and +AT morphology for clauses that were low in transitivity, with individuation of the patient as the main parameter for determining high transitivity. This transitivity marking is used today in both Tagalog and Toba Batak as a means of signalling the salience of the referents of patient NPs in discourse.

References

Fox, Barbara. 1982. "Clause Linking and Focus Systems in Some Western Austronesian Languages". Unpublished paper.

Givon, Talmy. 1979. On Understanding Grammar . New York:Academic Press.

Hopper, Paul J. 1979a. "Some Discourse Sources of Ergativity". Working Papers in Linguistics, Vol 11, No. 2. Honolulu: Department of Linguistics, University of Hawaii.

Hopper, Paul J. 1979b. "Aspect and Foregrounding in Discourse". Syntax and Semantics 12: Discourse and Syntax . Ed. by Talmy Givon. New York: Academic Press.

Hopper, Paul J. & Sandra A. Thompson. 1980. "Transitivity in Grammar and Discourse". Language, Vol 56, No. 2.

Thompson, Sandra A. Forthcoming. "The Passive in English: A Discourse Perspective." To appear in A Festschrift for Ilse Lehiste ed. by Robert Channon and Linda Shockey.