

An Experimental Study of Indonesian Voice

Andrew J. Gallagher
University of Oregon

1.0 Introduction

The experiment described in this paper is an attempt to resolve two complex issues of discourse analysis within a general theoretical framework of functional grammar. One task of analysts of discourse is to determine how mental constructs are coded or mapped onto grammatical form in order to communicate some set of propositions. This has led researchers to engage in attempts to determine the theme or topic of a discourse, two notions notoriously difficult to define. At the discourse level, it is tempting to determine the "aboutness" of the discourse, and it seems that everyone from the Prague School onwards has succumbed to this temptation, with very limited success. In fact, it has proved impossible to define these terms to everyone's satisfaction.

A similar problem has plagued the study of Indonesian discourse. Indonesian can be described as having two voices, active and passive marked by the verbal prefixes meN- and di- respectively. The di- prefix is only a part of a split person passive system, in which di- is used for 3rd person passive constructions while 1st and 2nd person constructions rely on word order and unmarked verb roots to indicate the passive voice (Moeliono 1988; Siu 1976; Wolff 1986a, 1986b). This paradigm is not without controversy however, since some Indonesianists ally Indonesian with other Austronesian languages using a focus rather than voice system (Dardjowidjojo 1974; Poedjosoedarmo 1985), or at least sufficiently unlike the Indo-European paradigm to be labelled as other than "passive" (Oglobin 1983; Verhaar 1976). Some have considered Indonesian to have a backgrounding and foregrounding system (Hopper 1979) or to exhibit syntactic or discourse ergativity. Many of these claims have gone unproven due to the lack of case marking on the NPs which

leads to inconclusive analyses, a matter summarized by Cumming and Wouk (1987). Other summaries of discourse-related issues can be found in works of Kaswanti Purwo (1986, 1988). These controversies notwithstanding, for the purposes of this paper, the terms active and passive will be used for simplicity.

Thus, in discourse analysis of Indonesian we are left with the task of determining the function of the meN- and di- prefixes. If they are some type of markers of voice alternations we still must explain the reasons for these alternations in real-world language use. One must consider the grammatical relations, the semantic roles and the pragmatic motivations involved. An alternate view of language use in general is to consider that the tired notions of theme or topic are, in fact, heuristic labels used to describe what is happening in the cognitive processing system of the speaker. Tomlin (1994) after years of experimental studies has taken a couple of steps back from this approach and uses a finer grain model of language production. Tomlin (to appear) argues convincingly that the notions of theme or topic are superfluous.

In this model there is no need for a linguistic category of theme or topic. Instead the grammar merely looks at the event representation directly and maps the current attentional focus onto subject. Since attentional focus is needed anyway, the overall grammar can be kept simple, and the concept of theme or topic is rendered superfluous to a theory of language or language production.

This was one of the motivations in designing the present experiment. Instead of struggling to describe and define a topic on a global discourse level, we have attempted to describe what is happening at the level of cognition. Analyzing the notion of voice in linguistics is one way of studying cognitive processes. Voice, as defined by Crystal (1985:329) is "a category used in the grammatical description of sentence or clause structure, primarily with reference to

verbs, to express the way sentences may alter the relationship between subject and object to the verb, without changing the meaning of the sentence." In common usage, in an active voice construction such as John ate the apple, the grammatical or syntactic subject John is also the actor or agent, while in a passive voice construction such as The apple was eaten by John, the syntactic subject apple is also the patient, i.e. that which is acted upon by the agent John. It is fairly easy to describe the movement of agent and patient to different positions in a clause or sentence, but it is another matter to explain in a real-world text or conversation, why one voice construction was chosen over the other.

2.0 Early Experiments

Studies have shown a link between voice and discourse and there have been attempts to determine the factors in the selection of voice, primarily by researchers in the tradition of psychology and cognition experiments. Johnson-Laird, as early as 1968, performed such experiments and found that "the **situation** of a communication (its socio-physical setting and linguistic context) probably exerts a decisive influence upon the form it takes and the way it is understood" (Johnson-Laird 1968:8). He also noted that active and passive voice constructions differ in their emphasis on the syntactic subject, determined by the word order. i.e. whether the agent or patient is in subject position. Turner and Rommetveit (1968:548) in an early study of active and passive sentences, used pictures to prime subjects to elicit active or passive voice, found that "the most salient semantic element, whether the actor of [sic] the acted-upon element, tended to become the subject of the sentence." Similarly, Tomlin (1994:528) reports that Flores d'Arcais (1975) in a study of Italian, used word cues to prime agent and patient, and found that "priming the agent led to actives 77% of the time; priming the patient led to passives 67% of the time."

For the present experiment I used a computer animated film designed by Tomlin (1994) involving colored fish

engaging in simple events. The fish were given one of five colors (black, white, red, green and yellow) but drawn in exactly the same shape and style. In the original pilot experiment, the fish participated solely in eating events, i.e. only two fish were visible on the computer screen at one time, and one fish inevitably ate the other. Just prior to the eating event, an arrow would appear on the screen pointing to one of the two fish. The subjects were instructed to keep their eye on the fish with the arrow, which in effect would allow a manipulation of their focal attention on a particular fish. The phrase attention detection is used to indicate that which is given attention at particular moments of utterance production. The subjects were also instructed to describe the event while it **happens**, in what is referred to as on-line description, in order to prevent the subject from having time to think and possibly reformulate the utterance with a newly attended referent. This also neatly reduces the effects of memory storage and retrieval on the language production (Cowan 1988, 1993).

Following previous voice studies, Tomlin (1994) investigated the production of active and passive voice constructions. Since the action of eating requires a transitive verb in English, Tomlin predicted that if the agent were primed, i.e. the fish that would eat is given the arrow, the subject would use active voice to describe the event. On the other hand, if the patient were primed, i.e. the fish that would be **eaten** is given the arrow, the subject would use the passive voice to describe the event. In the original study, Tomlin performed the experiment with speakers of eight different languages. The general finding was that languages grammatically code the focally attended referent in different ways. In English, Burmese and Indonesian, the focally attended referent is coded as the syntactic subject, i.e. the priming of agent resulted the active voice, and the priming of the patient resulted in passive voice, both more often than chance. However, in Mandarin, Polish, Russian, Bulgarian and Akan, there seems to be no grammatical coding of a focally attended referent. In some cases, agentivity, for instance, seems to override any other coding strategy. See

Tomlin (1994:544) for more details on these other languages.

3.0 The Indonesian Experiments

Before the present study of Indonesian, the original experimental design was revised several times. In the original experiment there were 31 consecutive eating events and Tomlin and his colleagues wondered if the subjects might fall into routines of using one voice construction or the other since the eating events were repetitive. In a new version, they added numerous filler trials of the fish participating in other actions. Thus to provide more variety, the fish were seen swimming past each other, above or below the other, kissing, following, and chasing each other. The resulting film contained 108 trials, lasting approximately 20 minutes.

The presentation of events was very carefully controlled and counterbalanced for fish color, speed, direction and the number of each type of event, to avoid any possible pattern to develop in the mind of the speaker. For instance, if the red fish always comes from the left, the speaker may notice this and focus his or her attention on this feature instead of attending to the arrow when it appears. In some of the previous data, the speakers seemed to be rushed in describing the event in the short time between the appearance of the arrow and the eating event. In the new version used in this study, there were two arrows used. One appeared at the time of approximately 2300 milliseconds before the eating event and lasted until approximately 1600ms before the fish was eaten. The second arrow then appeared 660ms before the eating event and stayed on the screen until the fish was fully consumed.

For the original pilot experiment in Indonesian, a single subject performed 100% according to prediction. When the agent was primed, the subject always used the meN- active voice construction, and when the patient was primed, the subject always used the di- passive voice construction. For the present study, seven subjects were used in the experiment with the revised film. All of the subjects were members of an introductory linguistics class and were given a form of extra

credit towards their final grade if they participated in the experiment. All were native speakers of Indonesian from Jakarta. All were undergraduate students in their 20s, six male and one female. All 108 trials were recorded on audio tape but only the target eating event descriptions were considered for data analysis.

Again, the on-line description task was designed to reduce the number of variables involved in discourse production and provide us with a fairly exact description of how the mental representation of the speaker will be coded onto grammatical form. The subjects heard a pre-recorded message at the beginning of the experiment and twice more after two short breaks that reminded them to "describe the events as they happen, and keep your eye on the character the arrow points at."

The hypothesis for the Indonesian experiment was as follows: Indonesian native speakers will code attention detection by using the meN- form of the verb when agent is primed, and the di- form of the verb when the patient is primed.

Consider some sample data to illustrate typical results, shown here with the trial number, the color of the primed fish, and the semantic role it plays in the eating event.

<u>Trl</u>	<u>Prime</u>	<u>Role</u>	<u>Utterance</u>
016	Green	AGT	Ikan hijau me-makan ikan merah. fish green ACT-eat fish red 'The green fish ate the red fish.'
018	White	PAT	Ikan putih di-makan ikan kuning. fish white PASS-eat fish yellow 'The white fish was eaten by the yellow fish.'

021 White AGT Ikan kuning di-makan ikan putih.
 fish yellow PASS-eat fish white
 'The yellow fish was eaten by the
 white fish.'

For example, if the agent were primed, and the subject attended to the primed agent fish, we found utterances such as in Trial 016 by subject #4. The green fish was primed as agent and the subject used the active meN- prefix, Ikan hijau memakan ikan merah, 'the green fish ate the red fish.' In cases when the patient was primed and the subject attended to the primed patient fish, the subject produced examples such as in Trial 018 in which he used the passive di- prefix, Ikan putih dimakan ikan kuning, 'the white fish was eaten by the yellow fish.' These were considered "hits". However, in Trial 021, he produced an unexpected utterance; although the agent was primed he used the passive di- prefix instead of the expected meN- prefix. We would have predicted an utterance equivalent to 'the white fish ate the yellow fish'; instead we found 'the yellow fish was eaten by the white fish'. This was considered a "miss."

4.0 Results

The predicted data and the results of the experiment are listed below. Recall that only the descriptions of the eating events were analyzed. In all there were 20 such active events and 20 such passive events, counterbalanced and distributed randomly throughout the experiment. The table in 4.1 illustrates that we expected 100% use of meN- if the agent were primed, and 100% use of di- if the patient were primed.

4.1 Expected Results

	active <u>meN</u> -	passive <u>di</u> -
agent primed	20/20 = 100%	0/20 = 0%
patient primed	0/20 = 0%	20/20 = 100%

Unfortunately, the results were inconclusive. Only four of the seven subjects performed as predicted. For these four subjects, when the agent was primed, they used the meN-prefix on the verb an average of 92.5% of the time; when the patient was primed they used the di- prefix 97.5% of the time. Thus these four subjects coded the events 94% as we predicted. However, the other three subjects (2,3,6) seemed to use a preferred-voice strategy. Subject #2 used 80% passive clauses, S#4 and #6 used 100% active and 90% active clauses respectively.

Since only 4 out of 7 subjects performed as predicted and I was not able to explain the unexpected results, I recruited several more subjects to see if a more clear pattern of usage would emerge. I was only able to use four more subjects and the results seemed to get worse instead of better. Section 5 summarizes the results of seven subjects from the original study and four additional subjects.

5.0 Discussion

5.1 Totals of Subjects with Expected Results (6 of 11)

	active <u>meN</u> -	passive <u>di</u> -
agent primed	106/119 = 89%	13/119 = 11%
patient primed	8/120 = 6.7%	112/120 = 93.3%

Table 5.1 shows that only 6 of the 11 performed as predicted. For all of these 6, they used the meN- prefix 89% of the time when the agent was primed, and the di- prefix 93.3% of the time when the patient was primed. The total number of hits was an average of 91.2% coding of attention detection onto grammatical form.

5.2 Total Hits: $218/239 = 91.2\%$ Coding of Attention Detection.

However, there remains a large residue of data summarized in 5.3 showing the voice used by each subject who did not seem to code attention detection in this way. Subject #2 was unique in using mostly passive voice. The other subjects used mostly active voice. Two used active voice exclusively.

5.3 Voice Used by Each Subject with Unexpected Results:

Subj: 2. 80% Passive
 3. 100% Active
 6. 90% Active
 8. 100% Active
 11. 98% Active

These data perhaps lead to more questions than answers. We are left to form hypotheses of why 6 of the subjects performed as predicted and 5 did not. Section 5.4 lists two possible hypotheses to explain the expected results in terms of attention detection.

5.4 Hypotheses to Explain Expected Results:

1) The subjects attended to the priming arrows and, 2) the subject's idiolect, through a process of attention detection, grammatically coded attention to agent role with meN- and patient role with di-. Section 5.5 lists possible reasons why we found unexpected results.

5.5 Hypotheses to Explain Unexpected Results:

1) The subjects did not attend to priming arrows and/or,
2) the subject's idiolect does not grammatically code agent and patient. Instead he or she primarily chooses active or passive voice.

I also checked to see if these results may have been an artifact of the experimental design. I suspected that perhaps these subjects were interacting with the computer screen as if reading a document and following the action from top to bottom and left to right. However, after re-analyzing the data from the first experiment, I found that 82% of the misses of the original 4 subjects who otherwise performed as predicted, were in fact focusing on the fish swimming from right to left, just the opposite of my suspicion.

For these misses, a similar percentage 82% were passive voice. It is interesting that in these attention-neutral conditions when the arrow is apparently ignored, the passive voice seems to be preferred among these subjects. One subject seemed to disregard the priming arrows and used passive voice 80% of the time. Of course, there is no statistical validity to data from so few subjects. And there remain the other 4 Indonesians who chose active voice most if not all of the time. Clearly these subjects preferred to attend to the agent role regardless of the experimental priming.

Some of my colleagues suggested that there might be dialectal influences on the use of Indonesian since Indonesians routinely speak more than one language besides Indonesian. I sent my original seven subjects questionnaires to find out more about this. It turns out that the person who used 80% passive claims to speak English, Spanish, Sundanese, Javanese, Minangkabau and Betawi. My hopes at a possible solution were quickly dashed however as I found that another subject who used 100% active also spoke English, Sundanese, Javanese, Minangkabau and Betawi. When I re-checked their files I discovered that they were in fact brothers. Needless to say, I could not posit any strong correlation between dialect and performance in the experiment.

Unfortunately, it remains unclear as to why some subjects choose one voice or the other. For those who use predominantly active voice or predominantly passive voice, we have no clear confirmation that they are making a linguistic choice, or have simply failed to perform in the experiment as we instructed them. It is possible that the subjects were not attending to the arrows at all and resorted to some other linguistic strategy to complete the task. Fortunately, an eye-tracking device will soon be installed in the laboratory to determine the exact eye movements of future subjects.

Finally, while the results remain inconclusive, these data offer insights to further study of Indonesian voice and discourse summarized in 6.0.

6.0 Importance of Experiments to the Study of Indonesian Voice and Discourse:

1) There is preliminary evidence that Indonesian marks voice by coding utterances with attention to agent roles with meN- and patient roles with di- forms of the verb; and 2) this type of experimental work may aid in analyzing the use of meN- and di- in discourse contexts in terms of attention and cognitive processing.

References

- Cowan, Nelson. 1988. Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin* 104.163-191.
- Cowan, Nelson. 1993. Activation, attention and short-term memory. *Memory & Cognition* 21.162-167.
- Crystal, David. 1985. A dictionary of linguistics and phonetics. 2nd ed. Oxford: Basil Blackwell.
- Cumming, Susanna, and Fay Wouk. 1987. Is there 'discourse ergativity' in Austronesian languages? *Lingua* 71.271-296.

- Dardjowidjojo, Soenjono. 1983. Passives as reflections of thought: A case in Indonesian. Language, literature and society: Working papers 1973 conference of American council of teachers of uncommonly taught languages: Occasional papers No. 1. Center for SE Asian studies, ed. by David W. Dellinger. DeKalb: Northern Illinois University.
- Flores d'Arcais, G.B. 1975. Some perceptual determinants of sentence construction. Studies in perception, ed. by G.B. Flores d'Arcais. Milan: Martello.
- Hopper, Paul. 1979. Aspect and foregrounding in discourse. Discourse and syntax (Syntax and semantics 12), ed. by T. Givon, 213-224. New York: Academic Press.
- Johnson-Laird, P.N. 1968. The choice of the passive voice in a communicative task. British Journal of Psychology 59.7-15.
- Kaswanti Purwo, Bambang. 1986. Strategi pemilihan meN- dan di- di dalam wacana Bahasa Indonesia. Linguistik Indonesia 4.1-14.
- Kaswanti Purwo, Bambang. 1988. Voice in Indonesian: a discourse study. Passive and Voice, ed. by Masayoshi Shabatani, 195-241. Amsterdam: John Benjamins.
- Moeliono, Anton M. 1988. Tata bahasa baku bahasa Indonesia. Jakarta: Balai Pustaka.
- Oglobin, A.K. 1983. Penelitian katakerja transitif dalam Bahasa Indonesia. Linguistik Indonesia 1.46-53.
- Poedjosoedarmo, Soepomo. 1985. Oleh. Linguistik Indonesia 3.10-17.
- Siu, Li Chuan. 1976. Essentials of Indonesian grammar. Sydney: Pustaka Malindo Publications.
- Tomlin, Russell. 1994. Focal attention, voice and word order. Word order in discourse, ed. by M. Noonan and P. Downing, 521-558. Amsterdam: John Benjamins.

- Tomlin, Russell. to appear. Mapping conceptual representations into linguistic representations: The role of attention in grammar. Conceptual and linguistic representations, ed. by J. Nuyts and E. Pedersen. Cambridge: Cambridge UP.
- Turner, Elizabeth Ann, and Ragnar Rommetveit. 1968. Focus of attention in recall of active and passive sentences. *Journal of Verbal Learning and Behavior* 7.543-548.
- Verhaar, John W.M. 1976. Some notes on the verbal passive in Indonesian. *Irian* 5.79-107.
- Wolff, John U. 1986. Formal Indonesian. Ithaca, NY: Cornell University Southeast Asia Program.
- Wolff, John U., Dede Oetomo and Daniel Fietkiewicz. 1986. Beginning Indonesian through self-instruction. Ithaca, NY: Cornell University Southeast Asian Program.

