The impact of verb subcategorization frequency on Standard Indonesian passives

Bernard A.J. Jap[1] & Yu-Yin Hsu[2]

[1] *Department of Humanities, Language and Translation, School of Arts and Social Sciences, Hong Kong Metropolitan University*
[2] *Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University*

Verb subcategorization frequency, the tendency of verbs to appear within specific syntactic structures, significantly influences sentence processing (Boland, 1997; MacDonald et al., 1994). Verbs often show biases towards certain constructions, affecting comprehension ease (Gahl et al., 2003; Menn et al., 2003). In Standard Indonesian (SI) both active and passive constructions are common (Sneddon et al., 2012), unlike in English, where passive sentences are less frequent and non-canonical, and are often harder to process (Ferreira, 2003),. The current study builds upon previous research exploring verb bias effects in SI (Jap & Hsu, 2024), which found increased processing costs for active-biased verbs in passive sentences compared to passive-biased verbs. We extend this by comparing active-biased and passive-biased passives to active sentences to examine how verb subcategorization frequency modulates passive sentence processing in SI. We hypothesized that a mismatch between the verb's preferred frame and the sentence structure will trigger more processing difficulty. Thus, passive sentences with active-bias verbs (i.e., verbs more frequently used in active voice) will be more difficult to process than those with passive-bias verbs, with active sentences as the control.

Thirty-one native SI speakers participated in an EEG experiment, word-by-word via rapid serial visual presentation (500ms per word, 100ms ISI). Stimuli comprised 80 active sentences with neutral-bias verbs (e.g., (1)), 60 passive sentences with passive-bias verbs (e.g., (2)), and 60 passive sentences with active-bias verbs (e.g., (3)). Verb bias classifications were determined using frequency data from the Indonesian Leipzig Corpora Collection (Goldhahn et al., 2012, see Figure 1a for frequency ratings of each verb token).

1. Active (Neutral): The man sent a staff member home yesterday from his office. *(Pria itu memulangkan kemarin seorang staf dari kantornya.)*
2. Passive (Passive-bias): The man was dismissed yesterday by the woman due to his behavior. *(Pria itu diberhentikan kemarin oleh wanita itu karena perbuatannya.)*
3. Passive (Active-bias): The man was pushed yesterday by the woman. *(Pria itu didorong kemarin oleh wanita itu.)*

EEG data were recorded using a 64-channel system. Event-related potentials (ERPs) were time-locked to the verb onset, the post-verbal adverb, and the second noun phrase (NP2). Cluster-based permutation tests (Maris & Oostenveld, 2007) were used for statistical analysis.

No significant differences were found between passive sentences with passive-bias verbs and active sentences, which may indicate that processing was 'facilitated' when verb bias aligns with the syntactic structure. While there are no effects on the verb and NP2, a significant P600 effect (p < .01) was observed at the post-verbal area (adverb) in passive sentences with active-bias verbs compared to active sentences. This late positivity, which was significant in clusters from 600 to 1000ms and largest over centro-parietal electrodes, suggests increased processing difficulty arising from the mismatch between verb subcategorization preference and the passive construction (see Figure 1b for the topographical distribution). ERP waveforms at representative channels are shown in Figure 1c.

These findings demonstrate that verb subcategorization frequency modulates passive sentence processing in SI. The P600 effect for active-bias passives challenges canonicity accounts based solely on word order frequency (Ferreira, 2003), emphasizing the influence of both lexical and structural factors on comprehension (Christiansen & MacDonald, 2009; Verhagen et al., 2018). This suggests a tight coupling of lexical and syntactic information during sentence processing, providing cross-linguistic insights into how lexical biases interact with syntactic processing, especially in languages with frequent passive use. Furthermore, extending Jap & Hsu (2024), the processing cost for active-bias passives extends to active sentences, highlighting the influence of verb subcategorization frequency.
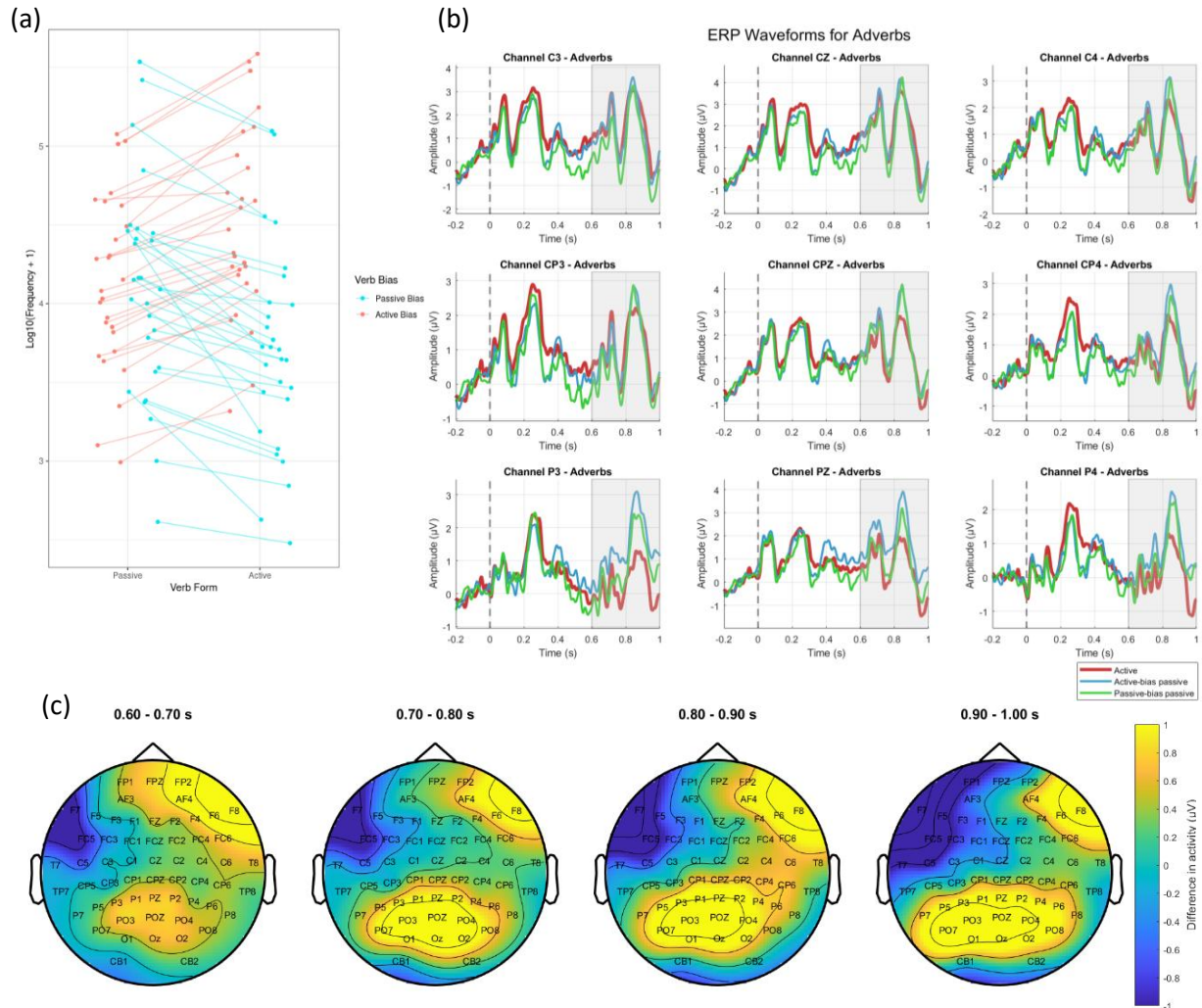
Figure 1. (a) Top left: Log10 transformed frequency of active and passive verb forms for each verb, categorized by verb bias (active or passive). (b) Top right: ERP waveforms for adverbs across nine electrode sites. The waveforms represent the average response for Active (red), Active-bias passive (blue), and Passive-bias passive (green) conditions. (c) Bottom: Topographic maps showing difference between Active-bias passive and Active conditions for adverbs across four time slices. From left to right: 600-700 ms, 700-800 ms, 800-900 ms, and 900-1000 ms.

References

Boland, J. E. (1997). The relationship between syntactic and semantic processes in sentence comprehension. *Language and Cognitive Processes*, *12*(4), 423–484.

Christiansen, M. H., & MacDonald, M. C. (2009). A usage-based approach to recursion in sentence processing. *Language Learning*, *59*(s1), 126–161.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*(2), 164–203.

Gahl, S., Menn, L., Ramsberger, G., Jurafsky, D. S., Elder, E., Rewega, M., & Holland, A. L. (2003). Syntactic frame and verb bias in aphasia: Plausibility judgments of undergoer-subject sentences. *Brain and Cognition*, *53*(2), 223–228.

Goldhahn, D., Eckart, T., & Quasthoff, U. (2012). Building large monolingual dictionaries at the Leipzig Corpora Collection: From 100 to 200 languages. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)* (pp. 759–765).

Jap, B. A. J., & Hsu, Y. Y. (2024). Does verb bias modulate thematic role assignment? Paper presented at *37th Annual Conference on Human Sentence Processing*.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676–703.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190.

Menn, L., Gahl, S., Holland, A. L., Ramsberger, G., & Jurafsky, D. S. (2003). Beyond canonical form: Verb-frame frequency affects verb production and comprehension. *Brain and Language*, *87*(1), 23–34.

Sneddon, J. N., Adelaar, A., Djenar, D. N., & Ewing, M. C. (2012). *Indonesian: A comprehensive grammar*. Routledge.

Verhagen, J., Heynderickx, M., & Pickering, M. J. (2018). Processing passive sentences: Evidence for a role of experience. *Journal of Memory and Language*, *99*, 119–132.