

Corpus-based approach meets LFG: Puzzling voice alternation in Indonesian

Pre-print (*under review*)

By:

Gede Primahadi Wijaya RAJEG; I Made RAJEG; I Wayan ARKA

2020

Based on [our talk](#) at the 25th *International Lexical-Functional Grammar*

Citation:

Rajeg, G. P. W., Rajeg, I. M., & Arka, I. W.. (2020). *Corpus-based approach meets LFG: Puzzling voice alternation in Indonesian* (Version 2). figshare.
<https://doi.org/10.6084/m9.figshare.12423788.v2>
([u'<https://doi.org/10.6084/m9.figshare.12311192.v1>', u'<https://lfg20.w.uib.no/program-june-23/>', u'<https://lfg20.w.uib.no/files/2020/05/21-Rajeg-Rajeg-Arka.pdf>', u'<https://youtu.be/BAVKyRA934E>']])

Abstract

This paper tests the meaning-preserving hypothesis of voice alternation in Indonesian. The hypothesis predicts that the meaning(s) encoded by a verb is available for both active and passive forms, differing only in the alignment of grammatical relations and semantic roles. Using quantitative corpus linguistic analysis, we argue that voice alternation needs to be relativised to (i) a certain sense of a verb and (ii) (statistical) usage constraints of the verb's semantics in certain voice. We also demonstrate the viability of LFG analytical machineries in capturing such empirical facts. Our study contributes to one of the big questions in language sciences, namely the relationship between form and meaning in language.

1 Introduction

This paper describes a novel approach to the study of grammatical voice (hereafter, voice) in Indonesian by providing fresh, corpus-based evidence that voice alternations for a given verb, especially between active and passive, are not always a meaning-preserving phenomenon (Kroeger 2005: 271) (cf. §2.3). The idea of “meaning-preserving” in active-passive alternation is that active and passive clauses, based on the same verb, should “describe the same kind of event” (Kroeger 2005: 271). Consider these examples in Indonesian.

- (1) *murid Go bie-pay yang meng-(k)ena-kan baju warna hitam.*
pupil NAME REL AV-hit-CAUS shirt colour black
'Go bie-pay's student who wears/puts on a black shirt.' (755227)¹
- (2) *Gaun yang di-kena-kan berwarna hitam*
dress REL PASS-hit-CAUS have.colour black
'The dress that is worn is black' (802596)

The base verb in (1) and (2) is *kena-kan* (derived from the verbal root *kena* 'hit; get into contact with') and conveys the event of 'wearing a clothing', both in its active form (hereafter AV) with *meN-* prefix (1) and in passive (hereafter PASS) with *di-* prefix (2). Looking at only these two sentences, the meaning-preserving status is indeed hold for voice alternation (of *kenakan* 'to wear' in AV and PASS). Based on quantitative corpus analyses (§3) on the usage of verbs derived from *kena* (see §2.2)), however, we argue for the following two empirical evidence:

- a. Certain sense of a given verb is significantly more frequently expressed in one voice type than in the other. This suggests that certain sense may be strongly and conventionally associated with certain form of voice compared to its voice-counterpart (Figure 2 and Figure 3).

¹ This number is sentence id in which the sentence is taken from the corpus. Unless otherwise indicated, all sentential examples are taken from one corpus file in the Indonesian Leipzig Corpora, namely *ind_mixed_2012_1M-sentences.txt*. We did not mention it in the text to save space. See §2.4 for further details on the corpus.

- b. Certain sense of a given verb can be directly constructed in one voice, in this case PASS, with no corresponding form in the other voice, in this case AV (§3.1). This indicates that (i) voice alternation should be relativised to certain sense of a verb such that one sense of a verb may not enter voice alternation, and that (ii) PASS form of a verb in a given sense is not always derived from its AV counterpart; hence not showing voice alternation, let alone a meaning-preserving property (Figure 2).

These two points touch on some of the key issues in language sciences. Two of these include (i) the questions of “how does form reflects meaning?” or “can we use difference in form as a measure of meaning?” (Janda 2016: 131); and (ii) the degree to which such form-meaning pairings form established units of the linguistic knowledge of a speaker (cf. Hilpert & Diessel 2016). To those ends, this paper is structured as follows. In §2, we present overview of related works this paper builds on (§2.1), followed by the studied verbs (§2.2), our aims (§2.3), and some methodological points (§2.4). Corpus analysis in §3 demonstrate the points in (a.) and (b.) above. Then, the proposed LFG analyses in §4 in order to capture the corpus findings include (i) argument-structure-based analyses of the entries of the morphological formatives and (ii) predicate composition and argument-fusion. We conclude in §5 with the implications of this study.

2 Corpus-based quantitative research on Indonesian voice

2.1 Some background

Corpus-based, quantitative research on voice in Austronesian languages focus on discourse-pragmatic factors (e.g. topicality of patient, transitivity of the event, grounding) that influence voice selection in discourse (among others, Wouk 1989, for Jakarta Indonesian; Pastika 1999, for Balinese; McDonnell 2016, for recent overview and his study in Besemah)². Despite the extensive research on voice, very little attention has been paid to the role of verb senses and their interaction with voice type of the verb. The question is whether voice alternation for a given verb stem interacts with the semantic potentials of the verb, given a verb can be polysemous.

McDonnell’s (2016: 242–244) *Collostructional Analysis* (Stefanowitsch 2013) on Besemah’s symmetrical voice constructions has shown that certain verbal roots more frequently occur in agentive voice than in patientive voice (see Gries & Stefanowitsch 2004 for a collostructional analysis for voice alternation in English). McDonnell further demonstrates that such degree of attraction plays a role in voice selection, in addition to the other factors (e.g. discourse transitivity and clausal mood). In line with our goal, McDonnell (2016: 250) notes an unexplored factor in voice selection, namely the possibility of semantic properties of the verbal root to account for strong preference of the verb to occur in AV or PASS construction (cf. the LFG

² Besemah is “a little-known Malayic language spoken in the remote highlands of South Sumatra in western Indonesia” (McDonnell 2016: 11).

analysis in §4.3). Another preliminary, quantitative study in Indonesian investigates the association between (metaphorical and literal) meaning and morphologically different verbs of the same root (based on *panas* ‘hot’) (that include voice morphologies) (Rajeg & Rajeg 2019). That study reveals that certain morphological form of a verb has stronger preference to be used in metaphorical context than in literal one (e.g., inceptive verb *memanas* ‘to become hot’ is significantly associated with metaphorical context, while passive causative *dipanaskan* ‘be heated up’ is significantly associated with literal context). We follow similar line of inquiry with other roots and focus.

2.2 Object of the present study

We start with derived verbs based on the verbal root *kena* ‘hit; get into contact with’. *Kena* is lexically Patient-oriented (cf. (3)), that is its syntactic subject (e.g. *orang* ‘person’ in (3)) is linked to Patient-like role. *Kena* is also associated with negative affectedness on the subject³.

- (3) *seperti orang yang kena hukuman di kursi listrik.*
 as.if person REL hit punishment at chair electricity
 ‘...as if a person who gets punished on an electrifying chair.’ (848667)

We studied derivatives of *kena* with two transitive suffixes, namely *-kan* (*kenakan*) and *-i* (*kenai*). These suffixes can express applicative/causative reading, depending on the roots (Arka et al. 2009). The stems *kenakan* and *kenai* can then take the *meN-* and *di-* prefixes that respectively encode AV and PASS voice types (see §4.1 for the lexical entries of the suffixes and voice prefixes). These two stems (*kenakan* and *kenai*) show puzzling behaviour in relation to their meanings as expressed in certain voice forms (key examples are in (4)a and (4)b).

- (4) a. *air kotor itu meng-(k)ena-i/*meng-(k)ena-kan baju Dimas.*
 water dirty DEM AV-hit-APPL/AV-hit-CAUS shirt NAME
 ‘... that dirty water hits/comes into touch with Dimas’ shirt.’ (774789)
- b. *motor kedua akan di-kena-i/di-kena-kan pajak sebesar 2 persen.*
 motor second FUT PASS-hit-APPL/-CAUS tax as.large 2 percent
 ‘...the second motorbike will be subject to/charged with 2% tax.’ (296558)

The original example in (4)a is with the *-i* form *mengenai* as the main predicate in AV, expressing the ‘physical hit or touch’ sense. We put the AV *-kan* form *mengenakan* to indicate that it cannot alternate with *mengenai* to convey the same ‘physical hit or touch’ sense. In contrast, in example (4)b, the *-kan* verbs *kenakan* can alternate with *kenai* in PASS in expressing

³ The ten most strongly attracted R1 collocates for *kena* (i.e. words immediately following *kena* within the sentence boundary) identified via Collostructional Analysis (Stefanowitsch 2013) are *pajak* ‘tax’, *batunya* ‘the stone’ (parts of idiom *kena batunya* ‘get into trouble’), *tipu* ‘deceive’, *marah* ‘angry/anger’, *racun* ‘poison’, *getahnya* ‘the resin’, *hukuman* ‘punishment’, *imbasnya* ‘the impact/effect’, *penyakit* ‘disease’, *semprot* ‘spray’ (which can have a metaphoric meaning of ‘getting a scolding’). They all evoke entities giving rise to negative affectedness on the subject of *kena*.

‘subject to’ sense. In other words, the PASS *dikenai* and *dikenakan* can equally express ‘subject to’ sense in (4)b. These are summarised in Figure 1.

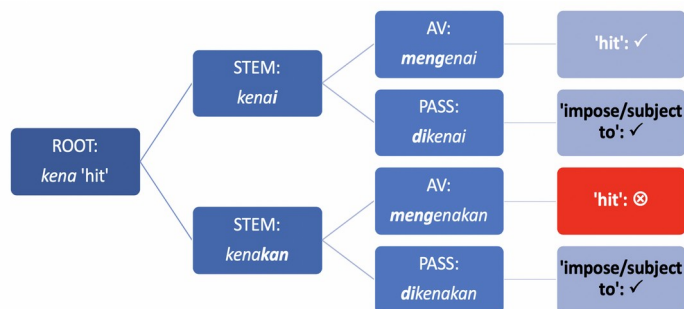


Figure 1: Form-meaning pairings for *kena*-derived verbs based on (4)a and (4)b.

2.3 Aims

The study consists of two parts: corpus analysis and LFG analysis. The corpus-based study seeks to determine whether different senses/meanings of verbal stems *kenakan* and *kenai* find similar distribution across AV and PASS (§3). Such an inquiry touches on, and would add quantitative nuance to, the meaning-preserving property of voice alternation for a given verb (see §1). The meaning-preserving hypothesis would predict that any senses expressed by a verbal stem in a given voice type (e.g. AV) are also available, or can be expressed in, PASS (cf. (1) and (2)). Even if we do find corpus attestation that each sense occurs in AV and PASS, the hypothesis would not predict whether certain sense is equally likely to be expressed in AV and PASS in language use (cf. McDonnell 2016: 243). In other words, meaning-preserving hypothesis does not predict the conventionality of certain sense to occur in certain voice type, given the sense is attested in AV and PASS. We scale up the amount of data we analysed because relying on a pair of examples (as in (1) and (2)) fail to capture asymmetry in the distribution of senses for a given verb in different voice types. LFG analysis (§4) will capture empirical fact about the dynamics of meaning construction and language use. It will analyse how semantic properties of the root *kena* join-forces with the semantics of voice morphology and valency-changing suffixes in the construction of meaning found in the derived verbs with *kena*.

2.4 Data source and coding

We retrieved all usage occurrences of the four target verbs, *mengenai* (N = 284 tokens), *dikenai* (N = 139), *mengenakan* (N = 1,101), and *dikenakan* (N = 446), from one corpus file, namely *ind_mixed_2012_1M-sentences.txt* (15,052,159 million word-tokens), a part of the *Indonesian Leipzig Corpora* collection (Quasthoff & Goldhahn 2013). This file consists mostly of shuffled sentences from Indonesian online news (Quasthoff & Goldhahn 2013: 26). The string *mengenai* actually occurs in total 7,148 tokens and

95.93% of which is its grammaticalised usage as a preposition meaning ‘concerning to’ (5) (Rajeg, Rajeg & Arka 2020: 336–339). The lexical usage of *mengenai* (as in (4)) was then manually identified.

- (5) *teman-temannya tahu mengenai siapa ‘kakaknya’ itu*
 friend.PL know concerning who older.sibling DEM
 ‘h(is/er) friends know *regarding* who h(is/er) older sibling is’ (212649)

The senses of each verb were coded based on two heuristic guidelines: (i) the description of the verb in the online *Kamus Besar Bahasa Indonesia* (KBBI) (the online Great Dictionary of Indonesian), and more importantly (ii) the semantic types of arguments co-occur with the verb. For instance, the ‘subject to/impose’ sense of *dikenakan* can be inferred from its co-occurrences with obligation-related arguments, such as *pajak* ‘tax’ in (4)b. Meanwhile, the ‘wear (clothing)’ sense of *dikenakan* is evoked when co-occurring with clothing-related arguments (see (1)). The primary ‘hit’ sense of *kenai* can be inferred when the event involves physical contact; examples (4)a in §2.2 and (6) below are the typical contexts. *Kenai* can also encode invisible/abstract affectedness, predominantly (i) medical affect, where a human or organ/parts of the body is affected by disease as in (7), and, (ii) to a small extent, psychological affect as in (8).

- (6) *orang yang di-kena-i anak panah itu terkapar mati*
 person REL PASS-hit-APPL child arrow DEM PASS.sprawled dead
 ‘...several people who got *hit* by those arrows were sprawled dead...’ (81198)
- (7) *Penyakit ini dapat meng-(k)ena-i pria dan wanita*
 disease DEM can AV-hit-APPL man and woman
 ‘This disease can *affect* (i.e. *hit*) men and woman ...’ (17661)
- (8) *tanggisan yang semata-mata meng-(k)ena-i pribadi debitur itu.*
 rebuttal REL merely AV-hit-APPL personality debtor DEM
 ‘a rebuttal that merely *affects* (i.e. *hit*) the personality of that debtor.’ (214779)

3 Corpus-based results

3.1 Senses for *kenai* in *PASS* and *AV*

The most frequent senses for *kenai* is ‘hit; contact; touch’ (N = 262; 61.94%), followed by ‘subject to/imposed’ (N = 124; 29.31%) and disease/mental ‘affect’ (N = 37; 8.75%). Figure 2 visualises the distribution of these senses in *PASS* and *AV* forms of *kenai*. The height of the bars represents percentages, with the raw numbers are given inside the bars.

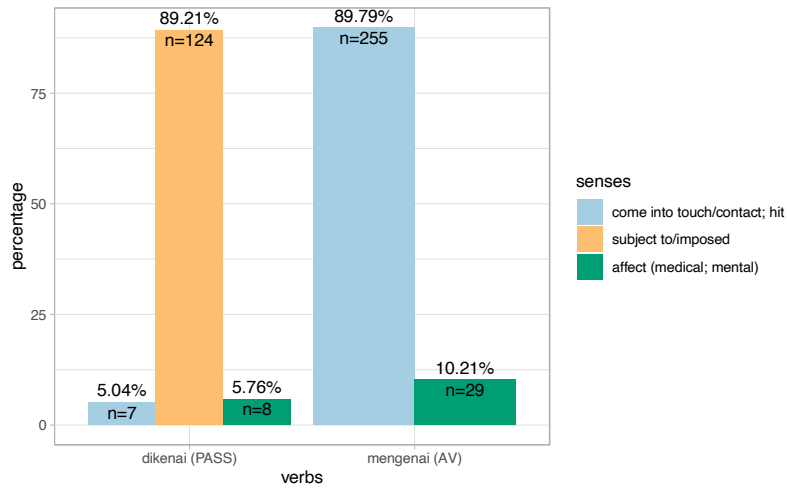


Figure 2: Distribution of senses for *kenai* in PASS and AV

It is clear that the distribution of senses for *kenai* is not equal across voice. The primary, physical sense of *kenai* ‘hit; come into touch/contact’ can indeed occur in PASS and AV (hence, categorically meaning preserving) but its proportion is much greater in AV (89.79 %) than in PASS (5.04%). Slight distributional difference could also be seen in ‘medical/mental affect’ sense. Figure 2 also provides an empirical evidence that voice alternation is not always meaning preserving. It is shown by the absence of ‘subject to/be imposed’ sense in AV; this sense is only found in PASS and intriguingly is the most frequent of all senses of *kenai* in PASS. This indicates that ‘subject to/be imposed’ is directly constructed and conventionalised in PASS. This ‘subject to/be imposed’ sense should not be regarded as a derivative of an (imaginary, underlying) AV form, which is empirically not attested for this sense in the corpus. Meaning-preserving hypothesis in voice alternation needs to be (i) relativised in terms of particular sense(s) of a given verb, and (ii) viewed probabilistically, as also shown in previous works that take the discourse-pragmatic approach (cf. §2.1, and §3.2).

A chi-square test for independence suggests that the asymmetric distribution of senses for *kenai* in PASS and AV (i) is statistically significant (i.e. cannot be due to chance) ($\chi^2 = 363.699$, $df = 2$, $p_{\text{two-tailed}} < 0.001$) and (ii) demonstrates a highly strong effect size (Cramer’s $V = 0.927$)⁴. The effect is indicated by the significantly strong preference of ‘hit; come into touch/contact’ sense to be expressed in AV (i.e. it has positive residuals in AV)⁵ and of ‘subject to/be imposed’ sense in PASS without AV occurrence.

⁴ Cramer’s V is a measure of effect size that is independent of sample size, unlike the significance level, which is dependent on the sample size (Levshina 2015: 209). Cramer’s V ranges from 0 (no association) to 1 (strong and perfect association). Strong effect size is shown by Cramer’s V value equal to or greater than 5 (Levshina 2015:209).

3.2 Senses for *kenakan* in PASS and AV

The lion share of the usage occurrences of *kenakan* convey the ‘wear; put on’ sense (N = 1,182; 77.31%), followed by ‘subject to/imposed’ (N = 301; 19.69%) and other senses (N = 46; 3.01%). Figure 3 visualises the distribution of these senses in PASS and AV forms of *kenakan*.

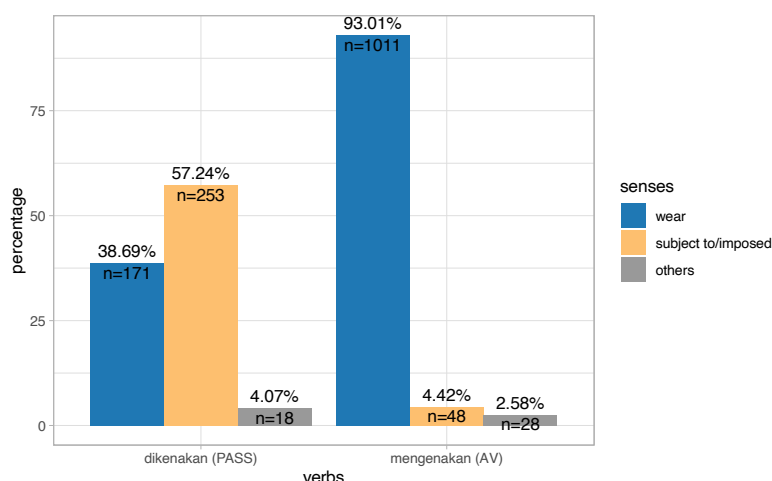


Figure 3: Distribution of senses for *kenakan* in PASS and AV

Similar trend of distributional asymmetry can be seen from Figure 3, as observed previously in Figure 2. The proportion of the two most frequent senses for *kenakan* (i.e. ‘wear’ and ‘subject to/be imposed’) differs in AV and PASS. ‘Wear; put on’, while indeed attested in AV and PASS, is much more frequently expressed in AV (93.01%) than in PASS (38.69%). In contrast, ‘subject to/be imposed’ is expressed much more frequently in PASS (57.24%) than in AV (4.42%). The chi-square test for independence suggests that this asymmetric distribution is statistically significant ($\chi^2 = 567.676$, $df = 2$, $p_{\text{two-tailed}} < 0.001$) and has a strong effect size (Cramer’s $V = 0.609$). The effect is shown by the significantly strong preference of ‘wear’ to be encoded in AV (but dispreferred in PASS) and of ‘subject to/imposed’ in PASS (but dispreferred in AV).

An important point to note on the PASS of the two stems *kenai* (§3.1) and *kenakan* is their similar semantic trait, namely predominantly conveying the ‘subject to/be imposed’ sense. This similarity unsurprisingly accounts for the fact that PASS *dikenai* and *dikenakan* can be interchangeably used to express ‘subject to/be imposed’ (in §2.2 example (4)b).

⁵ Space prevents us to include an Association plot in the manuscript, showing this strong preference effect for *kenai* as well as for *kenakan* (§3.2), but we are preparing the plot as part of open-access supplementary materials and will update this footnote with the DOI.

4 LFG Analysis

The LFG analysis consists of two components. The first one is an argument-structure based analysis, with entries of the morphological formatives: the root (*kena*), the transitiviser (*-i/-kan*), and the voice prefix (*meN-/di-*). The second component is principles for predicate composition, argument fusion and argument linking to capture (i) relative argument prominence, markedness, voice selection mechanism, (ii) constructional meaning in morphology and syntax, and (iii) restriction on semantic co-occurrences that evoke certain senses, and the AV/PASS preferential usage of these senses as reported in §3. Each of these components of analysis is discussed in order.

4.1 Lexical entry, argument structure and prominence.

We adopt a traditional morpheme-based analysis of Indonesian morphology, where the affixes including the voice and the transitivisers *-i/-kan* have their entries. Space precludes a full discussion, but in this subsection we briefly outlines our simplified a-structure representation of the lexical entry that captures prominence⁶ in grammar; see Arka et al (2009) for details. For example, the verbal root *kena* and the semantically bleached out or grammaticalised form *mengenai* as a verbal preposition are represented as having compact lexical entries shown in (9).

- (9) a. *kena* V (↑PRED) = ‘hit<(a:t), ^P:g>’
 b. *mengenai* P (↑PRED) = ‘concerning to<OBJ:P>’

The root *kena* (9)a is a semi-transitive verb (V), carrying an optional non-prominent theme-like actor causer and an obligatory P:goal argument. The semantic macro-role is represented in lower case, e.g. <a>; meaning that the role is actor-like, and that it is not prominent. The role representation in upper case, e.g. <P>, means that the role is highly prominent, with ^P meaning top-most prominent P in the a-structure list. The role in lower case after double colons as in <P:g> represent a specific thematic role. Thus, <^P:g> means ‘the most prominent macro-role (P) which is thematically a goal’. Prominence is also determined by structural embedding; the matrix argument is more prominent than a subordinate argument. Assuming GF-linking principles further discussed below in §4.2, the most prominent argument is selected as SUBJ by default, unless otherwise marked differently. Thus, given the lexical entry of the root *kena* in (9)a, we can account for data points as in (10) where <P:g> is linked to SUBJ, and <a:t> is possibly absent (the agent in (10) could be understood as the possessor *mereka* ‘their’ in the noun phrase *panah mereka* ‘their arrow’).

⁶ Prominence here relates to the idea of argument-ranking, which can be based on three levels. First, surface grammatical relations (i.e. syntactically privileged): SUBJ-PIVOT>non SUBJ-PIVOT; CORE>Non-CORE. Second, semantic/thematic role: AGENT/ACTOR>Non-ACTOR (A > Ground > Theme). Third, discourse pragmatics: for instance, TOPIC > non-TOP (Arka 2017; Sells 2001: 360).

- (10) *Seorang* *sahabat* ***kena*** (*panah* *mereka*) *hingga* *tewas*.
 ART.INDEF friend hit arrow 3PL.POSS until dead
 ‘A friend got *hit* (with their arrow) until (s)he is dead.’ (194)

It should be noted that *kena* carries a complex of inter-related senses, schematised in Figure 4; ‘hit’ is the semantic core with its sub-senses, which can interact with the semantics of the morphological formatives (cf. §4.3).



Figure 4: Semantic network of *kena* ‘hit’

The entry for *mengenai* (9)b says that the word is a preposition albeit in verbal form. Crucially, it comes with a very specific meaning, ‘concerning to’. Morphologically its form remains verbally transparent, containing transitive AV elements (i.e. *meN-* + *kena* + *-i*). In its P function, the meaning has been over time bleached off (i.e. grammaticalised). The grammaticalisation of *mengenai* into a function word of connective also accounts for the high token-frequency of *mengenai* in this function in the corpus (i.e. 95.91% [n=6,856] of 7,148 tokens). Like any other entries of P in Indonesian, the P *mengenai* has a sole P-like argument, linked to OBJ represented as <OBJ:P>. As such, it does not allow a passive alternation. Furthermore, the P *mengenai* ‘concerning to’ competes with other *kena*-based forms, which compositionally have general compositional meanings potentially deriving such a meaning. However, arguably, under the Pāṇinian (or Elsewhere) Principle, the more specific (‘concerning to’) meaning wins out, blocking any other morphological structure for the same meaning. This could explain the fact that *mengenakan* cannot be used for ‘concerning to’.

Voice prefixes also carry their own lexical entries; the AV and PASS entries are given in (11)a and (11)b respectively. These affixes are analysed as carrying their own argument structures, and voice affixation involves predicate composition and argument fusion. The fusion of the matrix and embedded arguments of the stem’s predicate is indicated by the connecting lines. The effect of fusion captures the effect of voice alternation in terms of prominence alternation and SUBJ selection. That is, the AV results in the fusion of the stem’s actor argument with the matrix $\wedge A$, which is therefore selected as SUBJ. In contrast, PASS promotes the stem’s patient to $\wedge P$ and is therefore selectable as SUBJ. More specifications will be given to the entries

of these voice prefixes to capture the linking within ParGram-style OT analysis; further discussed in §4.2 below.

(11) a. *meN* PREF (\uparrow PRED) = ‘AV<^A, (p), ‘STEM_PRED<a, p>’>’

b. *di-* PREF (\uparrow PRED) = ‘PASS<^P, ‘STEM_PRED<a, p>’ | ()>’

The suffixes *-i/-kan* are polysemous; their affixation results in causativisation or applicativisation depending on the nature of the stem (Arka et al. 2009). They carry their own predicate argument structures. The entries in (12) represent general information of these transitivisers showing two important points. First, the two suffixes represent matrix predicated of AFFECT, capturing the highly salient semantic conceptual units of transitive events (Jackendoff 1990) in which A affects P resulting in some kind of change as depicted by the meaning of the stem/root.

(12) a. *-i* SUFF (\uparrow PRED) = ‘AFFECT <A, P:goal/loc, ‘STEM_PRED<__, ()>’>’

b. *-kan* SUFF (\uparrow PRED) = ‘AFFECT <A, P ‘STEM_PRED<__, ()>’>’

Second, the entries also show the main distinction between the two transitivisers. The suffix *-i* specifies that the fused patient-like arguments must be thematically associated goal/locative, capturing the locative applicative/causative function of *-i*. The suffix *-kan* has no such thematic restriction, accounting for its more general functions that include benefactive/instrumental/theme applicatives as well as general non-locative causatives. There is also some overlap as both involve patient-like argument fusion (Arka et al. 2009). As we shall see in §4.3, there are different fusion options both actor and patient arguments giving rise to different realisations of arguments of the stem *kena*.

4.2 *Markedness and voice selection*

We adopt an OT-like, constraint-based argument linking and voice selection principles as implemented in Parallel Grammar (ParGram)⁷. Argument roles are ranked in terms of their prominence as outlined in §4.1 (e.g. A>G>T). In addition, surface grammatical functions (GFs) are also ranked (e.g. SUBJ>OBJ>OBL) (Arka 2017). In the OT-based voice selection system, arguments compete for their SUBJ linking; each has its own constraint (see Sells 2001, and the references therein). Crucially, linking constraints are also ranked; each is violable to the satisfaction of higher constraints, to yield an optimal output. To account for the symmetrical voice system in Indonesian

⁷ <https://ling.sprachwiss.uni-konstanz.de/pages/xle/doc/xle.html>

(in which both AV and PASS voice selections are equally morphologically marked),⁸ we propose a ranking of voice selection constraints given in (13).

(13)

Mark_L	Mark_5	+Mark_5	Mark_4	Mark_3	Mark_2	Mark_1
$\wedge[P/G]_L = \text{SUB}$	$*\wedge \text{ARG} = \text{SUB} \ \& \ * \emptyset$	$*\wedge \text{ARG} = \text{SUB} \ \& \ * \emptyset$	$*\text{NCORE}/\text{SUB} \ \& \ * \emptyset$	$*\text{T} = \text{SUB} \ \& \ * \emptyset$	$*\text{P}/\text{G} = \text{SUB} \ \& \ * \emptyset$	$*\wedge \text{A} = \text{SUB}$

First, a lexically specified linking constraint labelled ‘Mark_L’ (placed at the left most column in (13) is the most prominent one. The representation $\wedge[P/G]_L = \text{SUB}$ (without a star (*)) means that a non-A argument can be linked to SUBJ without any marking. The idea of no marking needed is captured by the notation without “ \emptyset ”. Note the notation $*\emptyset$ is a negative constraint, meaning ‘Don’t link an ARG (role) to SUBJ without marking’. In short, $\wedge[P/G]_L = \text{SUB}$ accounts for output sentences of *kena* exemplified in (10) where the P argument of the root *kena* is linked to SUBJ, and its actor-like argument if present appears in a non-SUBJ function.

Second, other SUBJ selection constraints are stated as negative constraints. Mark_5 constraint of $[\wedge \text{ARG} = \text{SUB} \ \& \ * \emptyset]$ is satisfied in the AV structure in Indonesian when A is linked to SUBJ because such SUBJ linking is marked by *meN-*. We can show the linking constraints explicitly in the updated lexical entry of *meN-* shown in (14)a. Likewise, the passive *di-* has its updated entry shown in (14)b.

(14) a. *meN* PREF ($\uparrow \text{PRED}$) = ‘AV < $\wedge \text{A}$, (p), ‘STEM_PRED < a, p >’>’

($\uparrow \text{OT-ORDER}_{[A]}$) = {Mark_5}

($\uparrow \text{SUBJ}$) σ = $\uparrow \sigma \wedge \text{A}$

b. *di-* PREF ($\uparrow \text{PRED}$) = ‘PASS < $\wedge \text{P}$, ‘STEM_PRED < a, p >’ | ($__$)>’

($\uparrow \text{OT-ORDER}_{[P]}$) = {Mark_5}.

($\uparrow \text{SUBJ}$) σ = $\uparrow \sigma \wedge \text{P}$

Finally, the lowest-ranked negative constraint $[\wedge \text{A} = \text{SUB}]$ is to capture the general default linking when the verb has no voice marker, but the A argument is linked to SUBJ. This is the AV structure without the AV morphology as seen in colloquial Indonesian in (15).

(15) *Untung saya bawa tustel*
 lucky 1SG bring camera
 ‘Luckily I bring a camera’ (3774)

⁸ Here we do not discuss Undergoer voice (UV) (Arka 2017: 116–119).

4.3 The dynamics of meaning interaction: *-i* vs. *-kan*

We are now ready to account for the preferential usage of voice selection (PASS vs. AV) associated with certain senses of *kena*. We begin by outlining the dynamics of meaning interaction due to the morphological derivation. We demonstrate that the analysis can capture complex cases explicitly. This includes how senses carried by voice and transitivity markers interact possibly to construct new senses, which then impose collocational, hence meaning, constraints on the derived verbs.

4.3.1 Evaluative meaning of *-i* and *kan*

The two transitivity markers carry different evaluative meanings, arguably due to the different thematic roles associated with their P argument. As mentioned earlier, the P of the transitivity marker *-i* is semantically locative/goal. The locative/goal P is therefore conceptually the target (i.e. end point) of the impact denoted by the *-i* verb. This property appears to be responsible for the strong negative evaluative meaning associated with *-i*. Consequently *-i* is not used to construct the ‘wear’ sense (only expressed by *-kan*; see §below). The negative affectedness sense of *-i* is incompatible with the essential socio-cultural meaning of ‘wear’, typically for positive artistic body decoration or health in Indonesian (or any other) culture.

Unlike *-i* (which focuses on the goal/loc affectedness), *-kan* introduces and focuses on the displacement process associated with the <theme> role (cf. Arka et al. 2009; Kroeger 2007). This is clear in the instrumental applicative use of *-kan* where the instrument role must be understood as an entity undergoing some kind of motion (16):

- (16) *Hamid(...) hendak mem-(p)ukul-kan kayu ke moncong buaya*
NAME intend AV-hit-CAUS.LOC wood to mouth crocodile
‘Hamid (...) intends to *smash* the wood to the crocodile’s mouth’ (10274)

In addition, *-kan* is associated with neutral or positive evaluative meaning. For example, only *-kan* is used in the benefactive structure (i.e. with positive evaluative meaning), as shown by example (17).

- (17) *ia pernah mem-bawa-kan saya kaligrafi Arab*
3SG ever AV-bring-APPL 1SG calligraphy Arabic
‘He once *brought* me Arabic calligraphy.’ (524017)

In short, while having some overlap (discussed in §4.1), the *-i* and *-kan* suffixes have different semantics, arguably related to the difference in thematic focus (goal/loc vs. theme). The suffix *-i*, not *-kan*, is highly compatible with the negative core sense of *kena* ‘hit’; cf. Figure 4). Affixing *kena* with *-i* consequently augments the negative affectedness of the root *kena*. Evidence for this comes from attestation in corpus, with the high proportion of occurrences of *-i* verbs expressing negative impact of the type of the example given in §3.1. For this reason, we represent *-i* and *-kan* to carry different superscripts, AFFECT⁻ and AFFECT⁺ respectively.

4.3.2 PASS only/Dominant PASS: ‘imposed, subjected to’ sense

Recall that an important finding of our study is that AV-PASS voice alternations do not always preserve meaning, and that there is evidence that certain dominant senses of verbs derived from *kena* correlate with particular voice types. We have seen that the negative ‘imposed, subjected to’ of *kenai* sense is attested only in the PASS form (Figure 2). This sense is also statistically more significant in PASS than in AV for stem *kenakan* (Figure 3). The proposed LFG-based analysis for this phenomenon is as follows.

First, the lexical entry of *kena* given in (18) below is fully specified with a complete set of information that constrains its combination with *-i/-kan* which gives rise to certain senses. The entry comes with conditional ‘if-then’ constraint sets (indicated by \Rightarrow) when *kena* is affixed with *-i* (a), or when it is affixed with *-kan* (b). The affixation results in predicate composition; e.g. $F(\uparrow \text{PRED}), (\rightarrow_H \text{PRED}_i)$ in (a) means the PRED of *kena* composes with that of *-i*. In this predicate composition process of *kena*+*-i*, there are inter-related constraints imposed. These consist of the permitted option of argument fusion type of its P:goal/loc role (either type 4 or 5, further discussed below), argument linking (with the P:goal/loc collecting the OT-ORDER constraint of +Mark5, meaning that the P role is the most preferred SUBJ), and a collocation constraint (with the theme role *kena*+*-i* being an abstract and non-wearable entity).

- (18) *kena* V $(\uparrow \text{PRED}) = \text{'HIT} \langle (a:_{th}) \text{ P:}_{goal} \rangle$
 (a) $(\uparrow \text{TR.SUFF_FORM}) = i \Rightarrow$
 $F(\uparrow \text{PRED}), (\rightarrow_H \text{PRED}_i)$
 $(\uparrow \text{FUSE.ARG TYPE}) = \{4, 5\}; [p = \text{P:}_{goal/loc}]$
 $(\uparrow \text{OT-ORDER}_{[p:goal]}) = \{+Mark5\}$
 $(\uparrow \text{GF SEM.TYPE}_{[a:t]}) = c \{ \sim \text{WEARABLE}, -\text{CONCRETE} \}$
 $\sim(\uparrow \text{VOICE-TYPE}) = \text{AV} \mid$
 (b) $\{(\uparrow \text{TR.SUFF_FORM}) = \text{KAN} \Rightarrow$
 $F(\uparrow \text{PRED}), (\rightarrow_H \text{PRED}_{KAN})$
 $(\uparrow \text{GF SEM.TYPE}_{[a:t]}) = c \sim \text{WEARABLE} -\text{CONCRETE}$
 $(\uparrow \text{FUSE.ARG TYPE}) = \{2, 3\}; [p = \{ \text{P:}_{goal[a:t]} \}]$
 $(\uparrow \text{OT-ORDER}_{[p:goal]}) = \{Mark5\}$
 $\}.$

To illustrate how the constraints specified in (18) work, let us consider the derivation of *dikenai* with the relevant example shown in (4)b. We focus on the rather abstract sense ‘imposed/subject to’, which appears only in the PASS form, *dikenai*. The diagrammatic representation of the derivation of *dikenai* is given in Figure 5 below. It can be explained as follows.

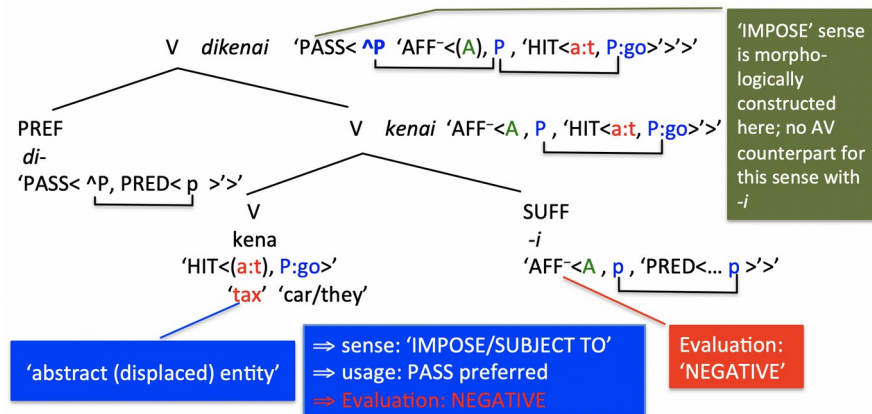


Figure 5: Derivation of *dikenai*

The predicate composition process of [*kena*+*i*] involves a single harmonious argument fusion: the matrix patient of *-i* is fused with the embedded P:goal. The single harmonious fusion is indicated by the roles having same blue colour, connected by a single line in the argument structure in Figure 5. This is Type 4 fusion, which results in a causative structure (see Arka et al. 2009, for other types of fusion).

The derivation by *-i* (which carries negative affectedness evaluation) augments the negative senses of the root *kena* ‘hit’ (e.g. ‘successfully affected’ and ‘(being) suffered’) (cf. Figure 4). These negative P-oriented senses are the main component senses of the ‘impose/subject to’ sense, which is only available in the PASS form. This provides evidence that such a meaning for *kenai* is constructed when it combines with the PASS formative *di-*. This is an instance of morphological construction (Booij 2010). It should be noted too that the ‘impose/subject to’ meaning also requires that the displaced entity involved (i.e. theme) belong to a particular semantic type, namely {~WEARABLE, -CONCRETE}. In example (4)b, this collocational constraint is satisfied by *pajak* ‘tax’.

The entry of *kena* (if it combines with *-i*) also imposes an OT-ORDER linking constraint, (\uparrow OT-ORDER_[P:goal])= {+Mark5}. That is, when *kena* is affixed with *-i*, it collects a high-ranked OT-mark, +Mark5, for its P:goal/loc. This means that, in the three-place predicate of *kenai* in which two non-actors available for SUBJ selection, the P:goal/loc is the most preferred SUBJ argument as it carries ‘+Mark5’. The acquisition of +Mark5 by P:goal/loc of *kenai* is an outcome of the augmentation of the negative affectedness P:goal/loc of *kena* by *-i*. This negative affectedness is further strengthened in the construction of the ‘impose/subject to’ meaning, only available in PASS with *kenai*. This strengthening of P-orientation is not surprising given the function of passive to focus on P, not A. The fact that the ‘impose/subject to’ meaning of *kenai* is available only in PASS is

captured by the negative constraint representation of $\sim(\uparrow\text{VOICE-TYPE})=\text{AV}$ in (18)a.

Turning to *-kena+kan*, we observe a slightly different pattern. As seen in §3.2, the negative ‘impose/subject to’ meaning is also available for *kenakan* in AV and PASS. Its occurrences in PASS are significantly more dominant than those in AV. It should be noted that the PASS of *kenakan* in this ‘impose/subject to’ meaning is significantly lower in frequency than the PASS of *kenai*. This low frequency suggests that *-kan*, in contrast to *-i*, is neutral in terms of its evaluation, and simply passes up the negative affectedness of the root *kena*. The fact that the AV/PASS alternation is allowed with *kenakan* is captured by having no $\sim(\uparrow\text{VOICE-TYPE})=\text{AV}$ in the entry of *kena* (18)b. In addition, the $(\uparrow\text{OT-ORDER}_{[\text{P:goal}]})=\{\text{Mark5}\}$ means that the P:goal/loc (like the other non-actor argument) is available (though not preferred) for SUBJ selection. Hence, our analysis allows an example like (19) where the theme, namely *bea keluar* ‘export fee’, is SUBJ:

- (19) *tidak ada bea keluar yang di-kena-kan bagi eksportir kakao.*
 NEG exist fee exit REL PASS-hit-CAUS for exporter cacao
 ‘there is no export fee that is *subjected to* cacao’s exporters’ (993142)

4.3.3 The ‘wear’ sense of *kenakan*

The ‘wear’ sense is only available for the composition of *kena* with *-kan*, not with *-i*. In addition, this sense is more dominant in AV than in PASS (§3.2). The relevant AV example has been shown in (1). To account for the derivation of *kenakan* ‘wear’ and its most preferred AV voice, the entry of *kena* needs to be updated. In addition to the conditional constraints (18)a-b, *kena* also has another set of constraints shown in (20)c. For simplicity, the constraints in (18)a-b are not repeated here. The diagrammatic representation of the derivation of AV *mengenakan* ‘wear’ is given in Figure 6, followed by some explanation below.

- (20) *kena* V $(\uparrow\text{PRED})=\text{‘HIT}<(\text{a:th}) \text{P:goal}>’$
 (c) $\{ (\uparrow\text{TR.SUFF_FORM})=\text{KAN} \Rightarrow$
 $\text{F}(\uparrow\text{PRED}), (\rightarrow_H \text{PRED}_{\text{KAN}})$
 $(\uparrow\text{FUSE.ARG TYPE})=\{4\}; [\text{A}=\text{P:goal}]$
 $(\uparrow\text{GF SEM.TYPE}_{[\text{a:t}]})=\text{c WEARABLE}$
 $(\uparrow\text{OT-ORDER}_{[\text{A}=\text{P:goal}]})=\{+\text{Mark5}\}$
 $(\uparrow\text{OT-ORDER}_{[\text{P}=\text{a:t}]})=\{\text{Mark5}\}$
 $\}$

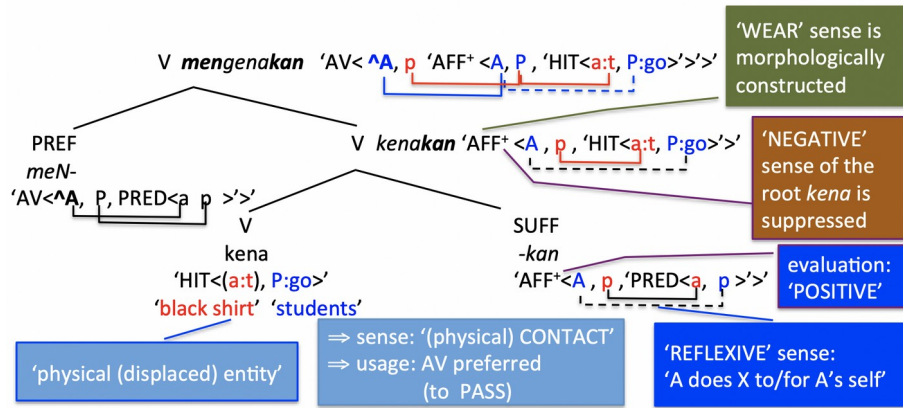


Figure 6: Derivation of *mengenakan*

The derivation diagram shows that the 'wear' sense is morphologically constructed at the [*kena*+*kan*] stem level. This is made possible by properties and constraints specified in the lexical entry of *kena* given in (20). Critical is the reflexive fusion (Type 4) required in the predication composition of [*kena*+*kan*], indicated by the dotted line connecting the matrix A of *-kan* and the embedded P:goal (in blue). The theme 'displaced' entity is also required to semantically belong to the WEARABLE type, satisfied by *baju warna hitam* 'black shirt' in (1). The sense of 'physical contact', central in the event conception of *kena* 'hit' is also salient; that is, the theme (i.e. shirt) ends up being located in A's own body (understood as P:goal). Importantly the matrix PRED *-kan* carries a positive evaluation (AFFECT+), suppressing the negative affectedness of the root *kena*. Moreover, the reflexive fusion of <A=P:goal> arguably augments the positive evaluation: one would not typically do something negative towards himself/herself, and in many cultures, people wear clothes or other accessories typically for positive social image, and/or other benefits such as healthy life. The lexical entry of *kena* (20) carries different OT-ORDER marks. {+Mark5} is specified for the A=P:goal role whereas {Mark5} (without a plus (+)) is for the theme argument. The difference is intended to capture the different voice-type distribution of 'wear' sense with AV preferred over PASS.

4.3.4 The Pāṇinian 'elsewhere' blocking effect

We have seen cases where certain senses/meanings are available or dominant in certain voice types; e.g. the 'impose/subject to' sense is only available in PASS (*dikenai*) for the stem *kenai*. For the stem *kenakan*, however, the 'impose/subject to' sense is available for both PASS and AV forms, but it is more dominant in PASS than in AV. These cases have been analysed as an outcome of complex OT-style prominence-based interaction of constraints and properties involved in predicate composition of the root *kena*, the transitiviser (*-i/-kan*) and the voice prefix (*meN/di-*). Overall the

interaction has given rise to a strong preference of a particular form-meaning pairing, which has the effect of blocking of the other logically form-meaning pairing. In what follows, we briefly discuss how such blocking can also be an outcome of the ‘elsewhere’ condition, also known as the Pāṇinian Determinism in constraint/rule interaction (Arregi & Nevins 2013).

The Pāṇinian determinism or ‘elsewhere’ condition says that a more specific rule or form-meaning pairing constraint has a priority over a more general one within the same paradigmatic domain. The more specific rule therefore blocks the more general one. In English verbal past tense paradigm, for example, the form-meaning pairing of *{went: {GO, PAST}}* is lexically specific; it blocks the application of the regular English past tense formation with the suffix *-ed*: **{[go+ed]: {GO, PAST}}*. In our present study on Indonesian, a blocking effect is seen in the expression of the core negative ‘physical contact/hit’ sense. As discussed earlier, this negative sense is expressed by the verbal root *kena* ‘hit’ (cf. entry (18)) and the transitiviser *-i* (cf. entry (12)). The set of forms possibly express this negative meaning are therefore *{kena, meng(k)enai, dikenai}*. Note that the derived verbs with *-kan*, such as *meng-(k)ena-kan* as shown in (4)a, are not possible because *-kan* carries a positive or neutral evaluative meaning.

Of particular interest is the fact that only the AV verb *meng(k)enai* (4)a, in addition to its root *kena* (cf. (3) and (10)a), can express this negative ‘physical contact/hit’ sense; the PASS form *di-kena-i* cannot. In the Pāṇinianism, we can say that the PASS *dikenai* is blocked by the lexically more specific form *kena*. That is, the root *kena* already specifically expresses the same passive-like meaning of negative ‘physical contact/hit’, which would also be expressed by the form *dikenai* generated by a more general PASS rule applied to the same root *kena*. Hence, the lexically specific rule of pairing of *{kena: ‘(negative) physical contact/hit’}* with the patient/goal linked to SUBJ takes precedence, and therefore blocks the PASS *dikenai*.

Our study also reveals an important finding to support the notion of morphological construction (Booij 2010): a particular sense is paired with (or constructed by) a specific morphological construction. Instances include the pairing of *{[meng+(k)ena+i]: ‘concern’}* and *{[di+kena+i]: ‘abstract impose/subject to’}*. The first one, as shown earlier, has undergone grammaticalisation into a preposition-like word (Rajeg, Rajeg & Arka 2020). The absence of the PASS *dikenai* to express this meaning of ‘concern’ appears to show another blocking effect, because the AV form *{[meng+(k)ena+i]: ‘concern’}* is morpho-constructionally specific (and fixed) for this form-meaning pairing such that a regular PASS is not possible to express the same meaning.

Likewise, the pairing of *{[di+kena+i]: ‘abstract impose/subject to’}* gives rise to the effect of blocking of its AV counterpart *[meng+(k)ena+i]*. That is, this abstract sense is morphologically constructed with the PASS *di-* being the essential part of the meaning construction. Thus, the word-level

construction of [*di+kena+i*] is specific for this meaning, blocking the AV counterpart with the same stem to express the same meaning.

However, the pairing of {[*di+kena+i*]: ‘abstract impose/subject to’} also competes with the verbs derived from [*kena+kan*] in the same semantic space of ‘abstract impose/subject to’ meaning. This fact highlights the well-known cross-linguistic pattern that there is no one-to-one pairing between form and meaning. Our statistical corpus-based evidence has revealed that the order of preference is *di+kena+i* in first place, followed by *di+kena+kan* and *meng+(k)ena+kan* in the second and third place respectively. This finding suggests that the graded preference of this kind can be thought of as an instance of ‘partial’ blocking. Issues of blocking in complex webs of form-meaning pairings across different paradigmatic domains appear to involve complex interactions of underlying constraints; this is an understudied area that needs further investigation involving more instances of verbal derivation. We leave this for future research.

5 Conclusion

The main goal of this paper was to test the meaning-preserving hypothesis in voice alternation (cf. §1 and §2.3). Using quantitative corpus linguistic analysis we argue that meaning-preserving hypothesis needs to be relativised to (i) the lexical meaning potential of the verbal stem in combination with voice morphologies (see the LFG analyses in §4.3), and (ii) (statistical) usage constraints of the verb’s semantics in certain voice (see §3). The basis of this argument is that a given verb can be polysemous where (i) a given sense of the verb can be significantly associated with one voice form than its voice-counterpart (cf. (a.) in §1), and (ii) certain sense for the same verb can be directly constructed in certain voice type ((b.) in §1), namely passive, without any evidence for the sense’s usage in active (hence, no evidence of voice alternation, let alone the meaning-preserving of that particular sense in different voice (cf. §3.1)). We also demonstrate that such empirical, quantitative findings on voice-meaning association can be captured using the constraint-based formalisms in LFG (i.e., lexical entry specification, predicate composition, argument-fusion, and OT-like constraints for voice selection). Moreover, the statistical preference that we report can also be framed with the classic idea of the *Elsewhere Principle* of blocking effect, going back to the 4th century BC Sanskrit grammarian, Pāṇini (§4.3.4).

Indeed, our conclusion is based on only one verbal root *kena*, with its derivation in different voice prefixes and two applicative/causative suffixes -*i* and -*kan* (cf. §2.2 and Figure 4). Be that as it may, our study supports few related works (see, in particular, McDonnell 2016; Gries & Stefanowitsch 2004), demonstrating statistical tendency of voice-specific, usage-preferences for a given verb(al root), and statistical association between certain meaning (i.e. metaphorical and literal) with certain voice morphologies (Rajeg & Rajeg 2019). Our usage-based, quantitative

approach adds nuance to the meaning-preserving hypothesis such that real usage preference is captured. This point is essential in usage-based linguistics that (i) considers the importance of frequency in the emergence, representation, and processing of linguistic units, and (ii) views linguistic knowledge varies along different continuum, one of them is conventionality and entrenchment (which could also be operationalised quantitatively (Hilpert & Diessel 2016)). The findings reported in this paper call into question the (implicitly presumed) equal status of PASS and AV alternation for a given verb stem, in terms of the conventionality and usage frequency in conveying certain sense in all voice types. Our study instead shows the asymmetry in the expression of meaning by a given voice form, hence touching on a big question of form-meaning relationship in language (Janda 2016: 131). Future experimental study is called for to assess how strong such statistical tendency is represented in the speakers' mind (i.e. do speakers also store in their linguistic repertoire such form-meaning pairing between a given voice form of a verb and its predominant meaning?).

6 References

- Arka, I Wayan. 2017. The core-oblique distinction in some Austronesian languages of Indonesia and beyond. *Linguistik Indonesia* 35(2). 101–144. <https://doi.org/10.26499/li.v35i2.58>.
- Arka, I Wayan, Mary Dalrymple, Meladel Mistica, Suriel Mofu, Avery D. Andrews & Jane Simpson. 2009. A linguistic and computational morphosyntactic analysis for the applicative *-i* in Indonesian. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG09 Conference*. CSLI Publications.
- Arregi, Karlos & Andrew Nevins. 2013. 12 Contextual Neutralization and the Elsewhere Principle. In Ora Matushansky & Alec P. Marantz (eds.), *Distributed Morphology today: Morphemes for Morris Halle*, 199–221. The MIT Press.
- Booij, Geert. 2010. Construction Morphology. *Language and Linguistics Compass* 4(7). 543–555. <https://doi.org/10.1111/j.1749-818X.2010.00213.x>.
- Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Extending collostructional analysis: A corpus-based perspective on “alternations.” *International Journal of Corpus Linguistics* 9(1). 97–129.
- Hilpert, Martin & Holger Diessel. 2016. Entrenchment in Construction Grammar. In Hans-Jörg Schmid (ed.), *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge*, 57–74. Berlin, Boston: De Gruyter. <https://doi.org/10.1515/9783110341423-004> (6 June, 2017).
- Jackendoff, Ray. 1990. *Semantic Structures*. Cambridge, Massachusetts: MIT Press.
- Janda, Laura A. 2016. Linguistic profiles: A quantitative approach to theoretical questions. *Język i metoda* 127–145.

- Kroeger, Paul. 2007. Morphosyntactic vs. morphosemantic functions of Indonesian *-kan*. In Annie Zaenen (ed.), *Architectures, Rules, and Preferences: Variations on Themes by Joan W. Bresnan*, 229–251. 1 edition. Stanford, Calif: Center for the Study of Language and Information.
- Kroeger, Paul R. 2005. *Analyzing Grammar: An Introduction*. Cambridge: Cambridge University Press.
- Levshina, Natalia. 2015. *How to do Linguistics with R: Data exploration and statistical analysis*. John Benjamins Publishing Company.
<https://doi.org/10.1075/z.195>.
- McDonnell, Bradley. 2016. *Symmetrical voice constructions in Besemah: A usage-based approach*. Santa Barbara, USA: University of California, Santa Barbara PhD dissertation.
<https://www.alexandria.ucsb.edu/lib/ark:/48907/f3mp53bw> (5 January, 2020).
- Pastika, I Wayan. 1999. *Voice selection in Balinese narrative discourse*. Canberra, A.C.T., Australia: The Australian National University PhD thesis.
- Quasthoff, Uwe & Dirk Goldhahn. 2013. *Indonesian corpora*. Leipzig, Germany: Abteilung Automatische Sprachverarbeitung, Institut für Informatik, Universität Leipzig. <http://asvdoku.informatik.uni-leipzig.de/corpora/data/uploads/corpus-building-vol7-ind.pdf> (27 July, 2015).
- Rajeg, Gede Primahadi Wijaya & I Made Rajeg. 2019. Pemahaman kuantitatif dasar dan penerapannya dalam mengkaji keterkaitan antara bentuk dan makna. *Linguistik Indonesia* 37(1). 13–31.
<https://doi.org/10.26180/5c6e1160b8d8a>.
- Rajeg, Gede Primahadi Wijaya, I Made Rajeg & I Wayan Arka. 2020. Contrasting the semantics of Indonesian *-kan* and *-i* verb pairs: A usage-based, constructional approach. In I Wayan Mulyawan, Made Sri Satyawati, I Nyoman Suparwa, Ketut Artawa & Maria Matildis Banda (eds.), *Prosiding Seminar Nasional Bahasa Ibu XII*, 328–344. Denpasar, Bali, Indonesia: Udayana University Press.
<https://doi.org/10.6084/m9.figshare.12311192>.
- Sells, Peter. 2001. Form and function in the typology of grammatical voice systems. In Géraldine Legendre, Jane Grimshaw & S. Vinker (eds.), *Optimality-Theoretic Syntax*, 355–391. Cambridge, Mass.: MIT Press.
- Stefanowitsch, Anatol. 2013. Collostructional analysis. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford handbook of Construction Grammar* (Oxford Handbooks Online), 290–306. Oxford: Oxford University Press.
<https://doi.org/10.1093/oxfordhb/9780195396683.013.0016>.
- Wouk, Fay. 1989. *The impact of discourse on grammar: Verb morphology in spoken Jakarta Indonesian*. United States -- California: University of California, Los Angeles Ph.D. (12 October, 2018).