

Corpus-based approach meets LFG: the puzzling case of voice alternations of kena-verbs in Indonesian

Gede Primahadi W. Rajeg
Universitas Udayana

I Made Rajeg
Universitas Udayana

I Wayan Arka

Australian National University/Universitas Udayana

Proceedings of the LFG'20 Conference

On-Line

Miriam Butt, Ida Toivonen (Editors)


2020

CSLI Publications

pages 307–327

<http://csli-publications.stanford.edu/LFG/2020>

Keywords: Lexical-Functional Grammar, quantitative corpus linguistics, voice alternation, Indonesian, form-meaning pairing, morphological construction

Rajeg, Gede Primahadi W., Rajeg, I Made, & Arka, I Wayan. 2020. Corpus-based approach meets LFG: the puzzling case of voice alternations of kena-verbs in Indonesian. In Butt, Miriam, & Toivonen, Ida (Eds.), *Proceedings of the LFG'20 Conference, On-Line*, 307–327. Stanford, CA: CSLI Publications. 

Abstract

This paper discusses the meaning-preserving hypothesis of voice alternation in Indonesian from an LFG perspective. The hypothesis predicts that the meaning encoded by a transitive 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 voices. We also demonstrate the viability of the LFG framework and related analytical issues in capturing such empirical facts.

1 Introduction*

This paper describes a novel approach to the study of grammatical voice (hereafter, voice) in Indonesian by providing fresh, corpus-based evidence in support of the claim that voice alternations in a given verb, especially between active and passive, are not always a meaning-preserving phenomenon (Kroeger 2005: 271). It further demonstrates that a given sense of a verb can be bias to one voice type. The idea of “meaning-preserving” in the active-passive alternation is that active and passive clauses involving the same verb should “describe the same kind of event” (Kroeger 2005: 271). To illustrate this point simply, we focus on verbs derived from the root *kena* ‘be hit; get into contact with’,¹ exemplified in (1) and (2).²

- (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)³
- (2) *Gaun yang di-kena-kan berwarna hitam*
dress REL PASS-hit-CAUS have.colour black
‘The dress that is worn is black’ (802596)

* We would like to thank (i) the two anonymous reviewers for their constructive comments on the first draft of the paper, and (ii) Charbel El-Khaissi (Australian National University) for proofreading our paper. Any remaining errors and inconsistencies are ours.

¹ One reviewer asked about our choice of the verb HIT, which is said to have elastic meanings, including strong metaphorical tendency, its tendency to be a light verb and as part of serial verb constructions. This should not be an issue. *Kena* ‘be hit’ represents the prototypical transitive impact verb in Indonesian, expressed by different forms appearing in different voice types with core and extended metaphorical meanings. Our research reported here is part of a larger research covering other verbs (cf. Rajeg, Rajeg & Arka 2020c). Any research for the meaning-preserving hypothesis should in principle also cover a wide range of verb types, including those with elastic meanings.

² Abbreviation in the interlinear glossing: 1, 2, 3, first, second, third person; APPL applicative; ART article; AV active voice; CAUS causative; DEM demonstrative; FUT future marker; INDEF indefinite; LOC locative; PASS passive voice; PL plural; POSS possessive; REL relativiser; SG singular.

³ This number is the sentence ID which indicates where in the corpus the sentence is taken. See §2.3 for further details on the corpus.

The base verb *kena-kan* in (1) and (2) conveys the event of ‘wearing a piece of clothing’, both in its active form (hereafter AV) with the prefix *meN-* (1) and in the passive (PASS) with the prefix *di-* (2). Given these two sentences, the meaning-preserving status does indeed hold for the voice alternation of *kenakan* ‘to wear’ in AV and PASS. The meaning-preserving hypothesis would categorically (and implicitly) predict that any senses expressed by a verbal stem in AV can be expressed in PASS, but would not predict whether a certain sense is equally likely to be expressed in AV and PASS (cf. McDonnell 2016: 243). That is, the hypothesis does not predict the conventionality of a certain sense according to the voice type given that the same sense is attested in both AV and PASS. We scale up the amount of data we analysed because relying on a pair of examples fails to capture distributional asymmetry of senses for a given verb in different voice types (see Bernolet & Coleman 2016 for a similar study in Dutch Dative Alternation). Based on quantitative analyses (§3) on the usage of verbs derived from *kena*, we argue the following two empirical claims:

- a. A certain sense of a given *kena*-based verb is significantly more frequently expressed in one voice type than in the other. This suggests that a certain sense may be strongly and conventionally associated with a certain voice compared to its voice-counterpart (Figure 2 and Figure 3).
- b. A certain sense of a given verb can be directly constructed in one voice, in this case PASS with *kena* as shown in this study, with no corresponding form in the other voice, in this case AV (§3.1 and Footnote 6). This indicates that (i) voice alternation should be relativised to a certain sense of a verb such that one sense of a verb may not enter voice alternation, and that (ii) the 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).

To these ends, this paper is structured as follows. In §2, we present an overview of related works that this paper builds on (§2.1), followed by the studied verbs (§2.2) and some methodological points (§2.3). The corpus analysis in §3 demonstrates the points in (a) and (b) above. Then, the proposed LFG analyses of the corpus findings in §4 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 and pointers to a future investigation.

2 Corpus-based quantitative research on Indonesian voice

2.1 Some background

Corpus-based, quantitative research on voice in Austronesian languages focuses 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).⁴ Despite extensive research on voice, very little attention has been paid to the role of verb senses and their interaction with the 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* (Gries & Stefanowitsch 2004) 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 a 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 for semantic properties of the verbal root to account for voice selection (cf. §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 revealed that certain morphological forms of a verb display a stronger preference for metaphorical contexts than literal ones (e.g., inceptive verb *memanas* 'to become hot' is significantly associated with metaphorical contexts, while the passive causative *dipanaskan* 'be heated up' is significantly associated with literal contexts). We follow a similar line of inquiry with other roots in Indonesian (see also Rajeg, Rajeg & Arka 2020c).

2.2 *Object of the present study*

Of particular interest are the syntactic-semantic differences of the derivatives of *kena* with two transitive suffixes, namely *-kan* (*kenakan*) and *-i* (*kenai*). Let us start with the properties of the root *kena* 'be hit; get into contact with', exemplified in (3). The verb *kena* is lexically Patient-oriented; that is, its syntactic subject (e.g. *orang* 'person' in (3)) is linked to a Patient-like role.⁵ *Kena* is also associated with negative affectedness on the subject. These two properties appear to be critical in constructing the derived meaning, and therefore in constraining the voice alternation in *kena*-based verbs.

⁴ Besemah is "a little-known Malayic language spoken in the remote highlands of South Sumatra in western Indonesia" (McDonnell 2016: 11).

⁵ The ten most strongly attracted R1 collocates for *kena* (i.e. words immediately following *kena* within the sentence boundary) identified via Collostructional Analysis (Stefanowitsch & Gries 2003) 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' and *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*.

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

It should be noted that the predicate *kena* carries a complex set of inter-related senses, schematised in Figure 1; ‘(be.)hit’ is the semantic core with its sub-senses, which can interact with the semantics of the morphological formatives (cf. §4.3). For instance, the *-kan/-i* suffixes can express an applicative or causative reading, depending on the semantic transitivity of the root (Arka et al. 2009).



Figure 1: Semantic network of *kena* ‘hit’

The stems *kenakan* and *kenai* can then take the *meN-* and *di-* prefixes that respectively encode AV and PASS voice types. 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. *Pengusaha meng-(k)ena-kan/*meng-(k)ena-i pajak*
 entrepreneur AV-be hit-CAUS/AV-be hit-APPL tax
 ‘Entrepreneurs imposes/charges tax (to their consumers)...’ (754049)
- b. *motor kedua akan di-kena-kan/di-kena-i pajak sebesar 2 persen.*
 motor second FUT PASS-be hit-CAUS/-APPL tax as.large 2 percent
 ‘...the second motorbike will be subject to/charged with 2% tax.’ (296558)

The original example in (4)a with the *-kan* AV verb *mengenakan* expresses the meaning, ‘subject to/impose’, however the AV *-i* form *mengenai* is not an acceptable alternative to convey the same ‘subject to/impose’ sense. In contrast, example (4)b shows that the *-kan* verb *kenakan* can alternate with *kenai* in PASS to express the meaning, ‘subject to’. In other words, the PASS alternation (*dikenai* and *dikenakan*) allows synonymy in expressing ‘subject to’ in (4)b, but such synonymy is not possible and infelicitous in the AV (4)a. LFG analysis (§4) will capture the empirical facts about the dynamics of meaning construction and language use, particularly how semantic properties of *kena* join forces with the semantics of voice

morphology and valency-changing suffixes in the construction of meaning that is found in the derived verbs.

2.3 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 across a total of 7,148 tokens among which 95.93% occur as a preposition meaning ‘concerning to’ (5) (Rajeg, Rajeg & Arka 2020a: 336–339). Despite this grammaticalised usage, *mengenai* can still be used as a lexical transitive verb (cf. (7) and (8) below) and this use of *mengenai* was 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 that 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 (a piece of 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; see examples (4)a in §2.2 and (6) below for 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-be 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-be 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-be 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 sense for *kenai* is the literal ‘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 given inside the bars.

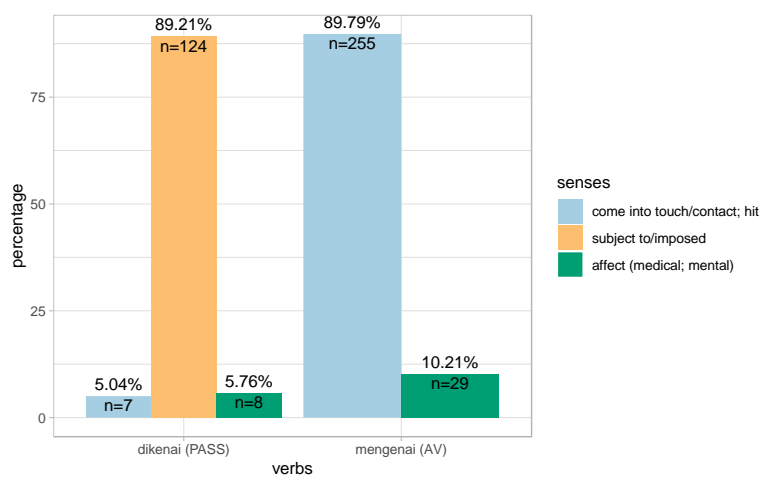


Figure 2: Distribution of senses for *kenai* in PASS and AV

It is clear that the distribution of senses for *kenai* is not equal across voice. The primary, physical sense ‘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%). A slight distributional difference can also be seen in the ‘medical/mental affect’ sense. Figure 2 also provides empirical evidence that voice alternation is not always meaning-preserving, but rather sensitive to a given sense of a verb. This is shown by the absence of the ‘subject to/be imposed’ sense in AV (cf. (4)a); this sense is only found in PASS and is the most frequent of all senses of *kenai* in PASS⁶. This indicates that ‘subject to/be imposed’ is directly

⁶ One reviewer asked how (any) corpus study can tell if the absence of AV *mengenai* to express ‘impose’ is a fact of grammar (a negative evidence), without recourse to native-speaker judgement in order to check if *mengenai* “could” mean ‘impose’, but would normally be infelicitous. We follow Stefanowitsch (2006; 2008), who proposes the corpus-based approach of negative evidence, and test for the statistical significance of a zero (0) frequency of ‘impose’ in AV *mengenai*. The goal is to check whether *mengenai* ‘impose’ is a possible form-meaning pair or not in Indonesian. Our analysis indicates that *mengenai* ‘impose’ is a highly significantly absent construction ($X^2 = 358.42$, $df = 1$, $p_{\text{two-tailed}} < 0.001$, $\phi = 0.921$), and our judgement as native speakers supports this corpus-based finding; the data and statistics are available at <http://bit.ly/negative-evidence>.

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.

A chi-square test for independence reveals that the asymmetric distribution of senses for *kenai* in PASS and AV (i) is statistically highly significant (i.e. cannot be due to chance) ($X^2 = 363.699$, $df = 2$, $p_{\text{two-tailed}} < 0.001$) and (ii) demonstrates a highly strong effect size (Cramer’s $V = 0.927$).⁷ The effect is indicated by the strong preference of the ‘hit; come into touch/contact’ sense expressed in AV (i.e. it has positive residuals⁸ in AV) and of the ‘subject to/be imposed’ sense in PASS without AV occurrence. Therefore, the meaning-preserving hypothesis in voice alternation needs to be (i) relativised in terms of particular sense(s) of a given verb (cf. Bernolet & Colleman 2016), and (ii) viewed probabilistically, as also shown in previous works from a discourse-pragmatic approach (cf. §2.1, and §3.2).

3.2 Senses for *kenakan* in PASS and AV

The lion share of *kenakan* occurrences 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 for *kenakan*.

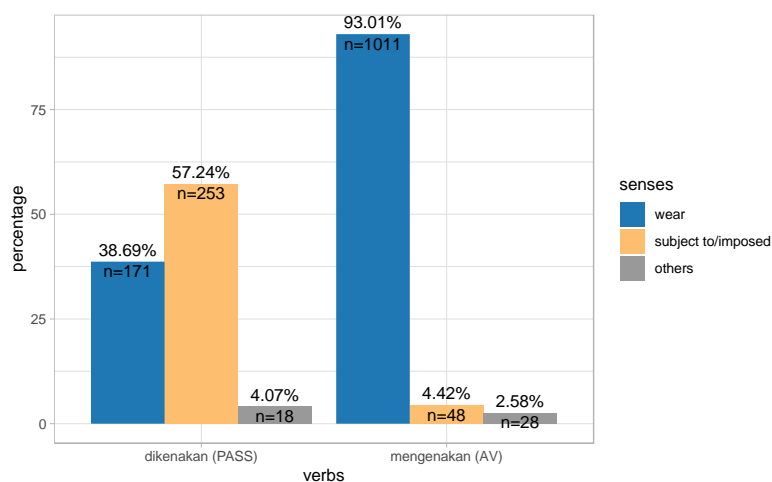


Figure 3: Distribution of senses for *kenakan* in PASS and AV

⁷ 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/large effect size is shown by Cramer’s V value equal to or greater than 0.5 (Levshina 2015: 209).

⁸ Space prevents us to include an Association plot that shows this strong preference effect for *kenai* and *kenakan* (§3.2), but see Rajeg, Rajeg & Arka (2020b) for links to the open-access supplementary materials, including the Association plot, data, and R codes.

A 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. While ‘wear; put on’ is indeed attested in AV and PASS, it 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 indicates that this asymmetric distribution is statistically highly significant ($X^2 = 567.676$, $df = 2$, $p_{\text{two-tailed}} < 0.001$) and has a strong effect size (Cramer’s $V = 0.609$). This effect is shown by the strong preference for ‘wear’ to be encoded in AV (but not preferred in PASS) and for ‘subject to/imposed’ in PASS (but not preferred in AV).

An important point in the PASS constructions involving the two stems *kenai* (§3.1) and *kenakan* is their similar semantic trait, which predominantly conveys 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’ (cf. example (4)b).

4 LFG Analysis

In this section we sketch out a LFG analysis, which 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 outlines principles for predicate composition, argument fusion and argument linking in order to capture, among other things, the voice selection mechanism, constructional meaning, the restriction on semantic co-occurrences that evokes 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-based linking.*

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 of the precise linking mechanism, but in this subsection we briefly outline our simplified a-structure representation of the lexical entry that captures prominence-based linking⁹ in grammar; see Arka et al (2009) for details. We adopt a version of a-structure-based linking as discussed in Arka (2003: 148–158), which is

⁹ 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 roles: AGENT/ACTOR > Non-ACTOR (A > Ground > Theme); agent > beneficiary > experiencer / goal > instrument > patient / theme > locative (Bresnan & Kanerva 1989; Butt 2014, among others). Third, discourse pragmatics: for instance, TOPIC > non-TOP (Arka 2017; Sells 2001: 360).

The suffixes *-i/-kan* carry their own predicate argument structures (Arka et al. 2009). The entries in (12) represent the general information of these transitivisers and demonstrate two important points. First, the two suffixes represent matrix predicates of AFFECT and capture the highly salient and conceptual semantic 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 < ARG₁, ARG₂, ‘STEM_PRED<__, ()>’>’
 (A) (P:goal/loc)
- b. *-kan* SUFF(\uparrow PRED) = ‘AFFECT < ARG₁, ARG₂, ‘STEM_PRED<__, ()>’>’
 (A) (P)

Second, the entries also show the main distinction between the two transitivisers. The suffix *-i* specifies that the fused patient-like (ARG₂) arguments must be associated to goal/locative roles, thus capturing the locative applicative/causative function of *-i*. The suffix *-kan* has no such thematic restriction, which accounts for its more general function including benefactive/instrumental/theme applicatives as well as general non-locative causatives. There is also some overlap between both suffixes as they involve patient-like argument fusion (Arka et al. 2009). As we shall see in §4.3, there are different fusion options for actor ARG₁ and patient ARG₂ arguments, which give rise to different realisations of arguments for the stem *kena*.

4.2 *Markedness and voice selection*

In the linking mechanism adopted here, arguments in the (syntacticised) argument structure are ranked in terms of their prominence, as outlined in §4.1 (e.g. actor ARG₁ outranks non-actor ARG₂). In addition, GFs are also ranked (e.g. SUBJ>OBJ>OBL) (Bresnan et al. 2015; Arka 2003). Arguments compete for their SUBJ linking; broadly speaking, the most prominent argument (typically actor ARG₁) is mapped onto SUBJ (see Arka 2003:151-156 for details). This linking is unmarked, in which case the (transitive) verb appears in its bare form. This can be seen in colloquial Indonesian in example (13), where the AV structure occurs without AV (SUBJ-selecting) morphology.

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

- (14) a. *Untung saya mem-bawa tustel*
 lucky 1SG AV-bring camera
 ‘Luckily I *bring* a camera’ (3774)

- b.
- | | | | | |
|-------------|--------|---------|---|---|
| <i>meN-</i> | PREFIX | (↑PRED) | = | ‘AV<ARG ₁ , ARG ₂ , ‘BRING< ARG ₁ , ARG ₂ >’>’ |
| | | | | <div style="display: flex; justify-content: center; align-items: center; gap: 20px;"> <div style="text-align: center;">SUBJ</div> <div style="text-align: center;">OBJ</div> </div> <div style="display: flex; justify-content: center; align-items: center; gap: 20px; margin-top: 5px;"> <div style="text-align: center;">(A)</div> <div style="text-align: center;">(P)</div> </div> |

However, Indonesian exhibits a symmetrical voice system in which both AV and PASS voice selections are equally morphologically marked (§1).¹¹ This voice symmetry is straightforwardly captured in our analysis by specifying that both AV and PASS prefixes have their own argument structures, as shown in (11). For example, the boldfaced verbal root in (13) can be morphologically marked for its AV type, as shown in (14)a. The argument structure of the verb *mem-bawa* ‘AV-bring’ is shown in (14)b. The AV marking results in the same linking as that in (13), in which the A ARG₁ and P ARG₂ are linked to SUBJ and OBJ respectively.

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 our analysis can capture complex cases explicitly. This includes how senses carried by voice and transitivity markers potentially interact to construct new senses, which then impose collocational constraints – and 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 goal-oriented. 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 (that is only expressed by *-kan*; see §4.3.3 below). The negative affectedness sense of *-i* is incompatible with the essential socio-cultural meaning of ‘wear’, which is typically used in Indonesian for positive, artistic body decoration.

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 (15):

¹¹ For simplicity, we do not discuss Undergoer voice (UV) (Arka 2017: 116–119) in this paper.

For ease of exposition - as the constraints apply and interact across different levels in the grammar in an intricate way - we formulate the set of constraints imposed in the predicate composition informally in the prose of (18) and (19) for *-i* and *-kan*, respectively. All of the constraints in (18) and (19) have a strong empirical basis (§3). They consist of a similar/overlapping and distinct set of constraints; the distinct ones are represented in bold: part (i) specifies argument fusion types at the level of argument structure, and part (ii) specifies the semantic nature of nominal types of ARG₂ at the level of semantic structure, both of which appear to constrain voice selections. Each part is further discussed briefly below, with reference to a specific example.

- (18) Constraints of PRED.COMPOSITION_I:
- i) Argument Fusion Type Constraint:
the goal ARG₁ of *kena* is fused with the **goal/locative** ARG₂ of the matrix PRED *-i* whereas **the displaced theme ARG₂ of *kena* can fuse with the matrix ARG₁**, or fuse with no matrix ARG, and it can constitute ARG₃ in the matrix argument structure of *-i*.
 - ii) Nominal Type Semantic and Voice/SUBJ-linking Constraints:
 - (a) If the displaced theme ARG₂ of *kena* is semantically ‘abstract’, and its goal ARG₁ is understood as **highly negatively affected**, then either ARG₁ or ARG₂ of *kena* is **obligatorily** linked to SUBJ (i.e. the matrix PRED is **obligatory in PASS**); or else,
 - (b) if the displaced ARG₂ of *kena* is fused with matrix ARG₁, and it is of the ‘concrete’ type, it also has to be of **the ‘non-wearable’ type**, and it is **highly preferred to be linked to SUBJ** with the matrix PRED appearing in AV.
- (19) Constraints of PRED.COMPOSITION_KAN:
- i) Argument Fusion Type Constraint:
the goal ARG₁ of *kena* is fused with the **patient** ARG₂ of the matrix PRED *-kan* whereas the **displaced theme ARG₂ of *kena* does not fuse with either matrix ARG₁ or ARG₂**; it constitutes ARG₃ in the matrix argument structure of *-kan*.
 - ii) Nominal Type Semantic and Voice/SUBJ-linking Constraints:
 - (a) If the displaced theme ARG₂ of *kena* is semantically ‘abstract’, then its realisation as SUBJ is **highly preferred to its realisation as non-SUBJ** (i.e. **the matrix PRED in PASS is not obligatory**); or else
 - (b) if ARG₂ is of the ‘concrete’ type, then it also has to be of **the ‘wearable’ type**, and it is **not preferred to be linked to SUBJ** as actor/goal ARG₁ is the preferred SUBJ with the matrix verb appearing in AV.

We are ready to discuss the derivation of the synonymous verbs *dikenai/dikenakan* ‘be subject to’ with their usage properties as attested in the corpus. The relevant example is (4)b, repeated in (20) with annotations of roles and GFs. The argument fusion of *dikenai* with the syntactic-

semantic properties of ‘be.imposed’ is given in (21). The subject ‘motor bike’ is ARG1 of *kena* and fused with ARG2 of *-i*, which is then selected as ARG1 by the passive *di-* and gives rise to the ‘be imposed/subject to’ sense. That is, this sense is constructed at the passive *di-* word level, which is an instance of morphological construction (Booij 2010). This word-level meaning construction is informally indicated by the horizontal curly bracket covering the whole morphological unit. The word-level constructed meaning of *dikenai* is semantically motivated by the highly-negative affectedness of the event (cf. line (18)ii.a, captured by AFFECT^{NEG} in (21)). Its construction is further motivated by the related semantic nominal type specific to *-i*, namely the theme ‘tax’ being something abstract/nonwearable. The agent ARG₁ of the stem *kenai* (i.e. the first argument in the inner argument structure list) is demoted and suppressed, indicated by a line connecting to ∅ in (21). While suppressed, its associated agentivity semantics (i.e. the event being volitionally imposed) remains.

- (20) *motor kedua akan di-kena-i/di-kena-kan pajak sebesar 2 persen.*
 motor second FUT PASS-be hit-APPL/-CAUS tax as.large 2 percent
 SUBJ:goal OBJ:theme
 ‘...the second motorbike will be *subject to/charged with 2% tax.*’ (296558)

- (21) SUBJ OBJ
di- ‘PASS< ARG₁, ‘AFFECT^{NEG} < ARG₁, ARG₂ ‘BE.HIT< ARG₁ ARG₂>’>’| ∅ >’
 (agent) (goal/loc) (goal) (theme)
 ‘motor bike’ ‘tax’
 ‘abstract’

The volitional ‘be.imposed’ sense is morphologically constructed at the level of the PASS formation *dikenai*, as no AV counterpart is possible.

Turning to *kena+kan*, we observe a slightly different pattern giving rise to a case of synonymy with the ‘impose/subject to’ sense as seen in (20). This is because the constraint of *-kan* in (18)ii.a overlaps with that of *-i* (19)ii.a. However, as seen in §3.2, the ‘impose/subject to’ sense for *kenakan* allows AV and PASS. Its occurrences in PASS are significantly more than those in AV. It should be noted that the proportion of PASS involving *kenakan* with this ‘impose/subject to’ meaning is lower than the PASS of *kenai*, suggesting that *-kan*, in contrast to *-i*, is neutral in terms of its affectedness evaluation. In other terms, *-kan* simply foregoes negative affectedness of the root *kena*.

- (22) SUBJ OBJ
di- ‘PASS< ARG₁, ‘AFFECT < ARG₁, ARG₂ ‘BE.HIT< ARG₁ ARG₂>’>’| ∅ >’
 (agent) (patient) (goal) (theme)
 ‘motor bike’ ‘tax’
 ‘abstract’

The volitional ‘impose/subject to’ sense is morphologically constructed at the level of the stem formation of *kenakan*, as the AV counterpart is possible.

The fact that the AV/PASS alternation is allowed with *kenakan* suggests that this ‘impose/subject to’ sense is constructed at the level of stem before voice morphology is added. Nevertheless, the stem still carries a prominent affected meaning because its PASS occurrences are more common than the AV counterparts. This empirical point is captured the ‘preference’ constraint when *kena* is affixed with *-kan*, as formulated in (19)ii.b. We do not attempt to formalise this preference constraint in this paper, but it can perhaps be done by incorporating ideas from Optimality Theory (see Sells 2001, and the references therein). The representation of the predicate composition in *di+kena+kan*, as shown in (22), is just like *di+kena+I*, as shown in (21). The only exception is that its AFFECT predicate is neutral (having no NEG superscript) and the volitional ‘impose’ meaning is constructed at the level of the the stem, which is denoted by the horizontal curly bracket partially covering the argument structure space.

4.3.3 The ‘wear’ sense of *kenakan*

The ‘wear’ sense is only available for the composition of *kena* with *-kan* and not with *-i*. In addition, this sense is more dominant in AV than in PASS (§3.2). The relevant AV example shown in (1) is repeated here in (23):

- (23) *murid Go bie-pay yang meng-(k)ena-kan baju warna hitam.*
 pupil NAME REL AV-hit-CAUS shirt colour black
 SUBJ:agent/goal OBJ:theme
 ‘Go bie-pay’s student who wears/puts on a black shirt.’ (755227)

The derivation and distribution of *kenakan* ‘wear’ with its preferred AV voice can be accounted for by the predicate composition constraints given in (19)i-ii.b. The AV *mengenakan* in sentence (23) can be analysed as having the predicate composition demonstrated in (24). The following points should be noted. First, the identified displaced theme ‘shirt’ meets the ‘concrete’/‘wearable’ requirement of the constraint, which triggers the preference for AV selection, as specified in (19)ii.b. The sense of ‘concrete’/‘physical contact’, which is central in the event conception of *kena* ‘hit’, is also salient; that is, the theme (i.e. shirt) ends up being located in the agent’s own body.

- (24) SUBJ OBJ
 meN- ‘AV< ARG₁, ARG₂ ‘AFFECT^{POS} < ARG₁, ARG₂ ‘BE.HIT< ARG₁ ARG₂>’>’>’
 (agent) (patient) (goal) (theme)
 ‘student’ ‘shirt’
 ‘concrete’
 ‘wearable’

The ‘wear’ sense is morphologically constructed at the level of the stem *kenakan* since AV/PASS counterpart is possible.

Second, the argument fusion shows harmonious fusion throughout the derivation processes, with higher arguments of the root and stem, ARG₁ and ARG₂, identified with matrix ARG₁ and ARG₂ respectively. This gives rise to a ‘reflexive meaning’ effect: the volitional agent (i.e. ARG₁) of *-kan*, which is also ARG₁ of AV, identified with the ‘student’ in example (23), is also the goal or locational target of the displaced theme ‘shirt’.

Third, the ‘wear’ sense is morphologically constructed at the [*kena+kan*] stem level, indicated by the horizontal curly bracket in (24). It allows AV/PASS alternation, with PASS permitted but not preferred (Figure 3).

4.3.4 *The preference constraint, morphological construction and the Pāṇinian ‘elsewhere’ blocking effect*

In this section, we address the issue of constraint interaction that was informally formulated in (18)-(19) and which specifies a ‘preference’ constraint to account for different kinds of ‘blocking’: strong and partial/weak blocking. We discuss the strong blocking in AV/PASS alternation, and relate it to the notion of morphological construction (Booij 2010) whereby a particular sense is paired with (or constructed by) a specific morphological pattern.

A clear blocking effect is observed in the case of verbs that display a very strong preference for a particular form-meaning pairing (e.g. *di+kena+i* ‘be.imposed’). This has the effect of blocking other logical form-meaning pairing (e.g. to express ‘impose’ in the AV form). In other words, while the Indonesian morphological derivation rule can produce AV/PASS forms *meng+(k)ena+i/di+kena+i*, the ‘impose’ sense with *kena+i* is strongly preferred in the PASS alternation, which blocks the AV alternation.

The strong preference constraint can also be understood as part of the broader constraint in rule competition, which is traditionally discussed under the rubric of the ‘elsewhere’ condition or Pāṇinian Determinism (Arregi & Nevins 2013). Such conditions state that a more specific rule or form-meaning pairing constraint in rule competition has a priority over a more general one within the same paradigmatic domain. The more specific rule therefore blocks the more general one. For example, the form-meaning pairing of {*went*: {GO, PAST}} in English is lexically specific; it blocks the application of the regular English past tense formation with the suffix *-ed*: *{[*go+ed*]: {GO, PAST}}. The non-existence of the form *meng+(k)ena+i* to express the ‘impose’ sense (in the AV form) can also be accounted for in terms of blocking with reference to specific morphological form-meaning pairing. That is, the form-meaning pairing of {*dikenai*: {IMPOSE, AFFECT^{NEG}, ABSTRACT.THEME}} is specific in expressing the ‘impose’

sense in its negatively affected meaning such that it blocks other forms from expressing the same meaning, including *meng-(k)enai*.

Also, of particular interest in the context of blocking is the fact that only the AV verb *meng(k)enai* (4)a, including its root *kena* (cf. (3) and (10)a), can express the negative ‘physical contact/hit’ sense; the PASS form *di-kena-i* cannot. Under Pāṇinian Determinism, *dikenai* is generated by a general PASS rule; it is blocked by the more lexically-specific form, *kena*. That is, the root *kena* specifically expresses the same passive-like meaning of negative ‘physically be.hit’.

Our study also reveals an instance of blocking that involves a grammaticalisation dimension in the pairing of {[*meng+(k)ena+i*]: ‘concern’}. The form-meaning pairing has undergone grammaticalisation into a preposition-like word (Rajeg, Rajeg & Arka 2020a). The absence of the PASS *dikenai* to express ‘concern’ can be thought of as a blocking effect because the AV form {[*meng+(k)ena+i*]: ‘concern’} is morpho-constructurally specific (and fixed) for this form-meaning pairing such that a regular PASS is unable to express the same meaning.

Turning to partial/weak blocking, we revisit the AV/PASS alternation in the stem *kenakan* with the ‘impose/subject to’ sense. This sense is available for both PASS and AV forms, but it is more predominant in PASS than in AV. We could say that PASS partially blocks AV. Furthermore, it should be noted that ‘impose’ is also expressed by [*di+kena+i*] and hence, (*di*)*kenai* also competes with verbs derived from [*kena+kan*] in the same semantic space of ‘impose’. These facts highlight 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. 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.

5 Conclusion

The main goal of this paper was to discuss the meaning-preserving hypothesis in voice alternation (cf. §1). Using quantitative corpus linguistic techniques, we argue that the 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 voices (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. point (a) in §1), and (ii) a certain sense for the same verb can be directly constructed in a certain voice type

(point (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 a different voice (cf. §3.1 and Footnote 6)). 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 preference constraints for voice selection). Moreover, the statistical preference that we report can also be framed within the classic idea of the *Elsewhere Principle* of blocking effect proposed by Pāṇini (§4.3.4), even though there remain issues of different degrees of (partial) blocking that need further analytical exploration.

Indeed, our conclusion is based on only one verbal root *kena*, with its derivation in different voice prefixes and two applicative/causative suffixes (i.e. *-i* and *-kan*). Be that as it may, our study supports few related works that demonstrate the statistical tendencies of voice-specific, usage-preferences for a given verb(al root) (see, in particular, McDonnell 2016; Gries & Stefanowitsch 2004), as well as the statistical association between certain senses and certain voice morphologies (Rajeg & Rajeg 2019; Rajeg, Rajeg & Arka 2020c; cf. Bernolet & Coleman 2016, for Dative Alternation). Our quantitative approach contributes nuance to the meaning-preserving hypothesis in such a way that real usage preference is captured. This point is essential in usage-based linguistics (Diessel 2017), which (i) considers the importance of frequency in the emergence, representation, and processing of linguistic units, and (ii) views linguistic knowledge as varying along different continuum, such as conventionality and entrenchment. Further study is needed to experimentally assess how strong the statistical tendency reported in this paper is represented in the speakers' mind: do speakers also store in their linguistic repertoire such form-meaning pairing between a given voice form of a (morphologically complex) verb and its predominant meaning? A related corpus-based and experimental study using sentence-production tasks in Indonesian CAUSED MOTION verbs reveals some convergence between participants' usage of the target verbs in certain voices and the corpus findings (Rajeg, Rajeg & Arka 2020c). This indicates that speakers may store statistical patterns of association between morphologically complex verbs and their predominant meanings. Our findings 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 senses in all voice types. We instead show the asymmetry in the expression of meaning by a given voice form.

6 References

- Arka, I Wayan. 2003. *Balinese morphosyntax: A lexical-functional approach*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University.
- 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*. Stanford: CSLI.
- Arka, I Wayan & Christopher D. Manning. 2008. Voice and grammatical relations in Indonesian: A new perspective. In Peter Austin & Simon Musgrave (eds.), *Voice and grammatical relations in Austronesian languages*, 45–69. Stanford: CSLI.
- 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.
- Bernolet, Sarah & Timothy Coleman. 2016. Sense-based and lexeme-based alternation biases in the Dutch dative alternation. In Jiyoung Yoon & Stefan Th. Gries (eds.), *Corpus-based approaches to Construction Grammar*, 165–198. Amsterdam: John Benjamins Publishing Company.
- Booij, Geert. 2010. *Construction Morphology*. Oxford: Oxford University Press.
- Bresnan, Joan, Ash Asudeh, Ida Toivonen & Stephen Wechsler. 2015. *Lexical-Functional Syntax*. Second edition. Malden, MA : Chichester, West Sussex: Wiley-Blackwell.
- Bresnan, Joan & Jonni Kanerva. 1989. Locative inversion in Chichewa: a case study of factorization in grammar. *Linguistic Inquiry* 20. 1–50.
- Butt, Miriam. 2006. *Theories of Case*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139164696>.
- Butt, Miriam. 2014. Control vs. complex predication: Identifying non-finite complements. *Natural Language & Linguistic Theory* 32(1). 165–190. <https://doi.org/10.1007/s11049-013-9217-5>.
- Diessel, Holger. 2017. Usage-based linguistics. (Ed.) Mark Aronoff. *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press.
- Falk, Yehuda. 2001. *Lexical-Functional Grammar*. Stanford: CSLI.
- Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Extending collocation analysis: A corpus-based perspective on “alternations.” *International Journal of Corpus Linguistics* 9(1). 97–129.
- Jackendoff, Ray. 1990. *Semantic Structures*. Cambridge, Massachusetts: MIT Press.
- Kroeger, Paul. 2005. *Analyzing Grammar: An Introduction*. Cambridge: Cambridge University Press.
- 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. Stanford: CSLI.

- Levshina, Natalia. 2015. *How to do Linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins Publishing Company.
- McDonnell, Bradley. 2016. *Symmetrical voice constructions in Besemah: A usage-based approach*. Santa Barbara, USA: University of California, Santa Barbara PhD dissertation.
- Pastika, I Wayan. 1999. *Voice selection in Balinese narrative discourse*. Canberra, Australia: The Australian National University PhD thesis.
- Quasthoff, Uwe & Dirk Goldhahn. 2013. *Indonesian corpora*. Leipzig, Germany: Abteilung Automatische Sprachverarbeitung, Institut für Informatik, Universität Leipzig.
- 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/10.26499/li.v37i1.87>.
- Rajeg, Gede Primahadi Wijaya, I Made Rajeg & I Wayan Arka. 2020a. 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>.
- Rajeg, Gede Primahadi Wijaya, I Made Rajeg & I Wayan Arka. 2020b. Supplementary materials for “Corpus-based approach meets LFG: Puzzling voice alternation in Indonesian.” Open Science Framework. <https://doi.org/10.17605/OSF.IO/YMD2V>.
- Rajeg, Gede Primahadi Wijaya, I Made Rajeg & I Wayan Arka. 2020c. Usage-based perspective on the meaning-preserving hypothesis in voice alternation: Corpus linguistic and experimental studies in Indonesian. Paper presented at the UCREL Corpus Research Seminar, Lancaster University, UK (Online). <https://doi.org/10.6084/m9.figshare.13283918.v1>.
- 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. 2006. Negative evidence and the raw frequency fallacy. *Corpus Linguistics and Linguistic Theory* 2(1). 61–77.
- Stefanowitsch, Anatol. 2008. Negative entrenchment: A usage-based approach to negative evidence. *Cognitive Linguistics* 19(3). 513–531. <https://doi.org/10.1515/COGL.2008.020>.
- Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2). 209–243.
- Wouk, Fay. 1989. *The impact of discourse on grammar: Verb morphology in spoken Jakarta Indonesian*. California: University of California, Los Angeles PhD dissertation.