I Made Rajeg*, Gede Primahadi Wijaya Rajeg and I Wayan Arka

Corpus linguistic and experimental studies on the meaning-preserving hypothesis in Indonesian voice alternations

https://doi.org/10.1515/lingvan-2020-0104
Received November 26, 2019; accepted June 1, 2021; published online January 28, 2022

Abstract: One essential feature of voice alternation is that active and passive clauses centred around a given verb express the same meaning: the "meaning-preserving" hypothesis. One effect of the alternation is the different linking of grammatical relations and semantic roles, which affects the identity of the subject. This paper investigates the meaning-preserving hypothesis in voice alternation in Indonesian from a quantitative usage-based perspective by combining corpus-based data with sentence-production experiment data. It analysed Indonesian CAUSED FORWARD/BACKWARD MOTION verbs and the distribution of their (non-)metaphoric senses in active and passive. The findings demonstrate the frequency effects and sense-sensitivity of voice alternation, such that a given voice type of a verb is strongly associated with certain senses. This finding provides initial support for a previous study on voice alternation in an Austronesian language, predicting that the verb's semantic properties may condition the statistical bias of the verb towards a particular voice. Some convergence between experimental and corpus findings indicates that participants demonstrate some representation of the strong association between a given voice form of the verb and the sense predominantly expressed in that form, highlighting the notion of item-specific representations of linguistic knowledge as found in construction grammar.

Keywords: construction grammar; Indonesian; metaphor; quantitative corpus linguistics; voice alternation

1 Introduction

One of the essential features of active (AV) and passive (PASS) voice alternation is that AV and PASS clauses centred around a given verb express the same meaning or sense: see (1) and (2).[1] This is referred to as the "meaning-preserving" hypothesis (Kroeger 2005: 271). The effects of voice alternation include the re-alignment of grammatical relations and semantic roles, which affects the identity of the subject, and the demotion of the AV agent in the PASS into an oblique or unexpressed argument (Kroeger 2004: 54).

(1) *Dindik* *Surabaya* ***me-maju-kan*** *jadwal* *ujian*
ABBREV NAME AV-move.forward-CAUS schedule exam
'The Education Department of Surabaya city *moved* the exam schedule *forward* …'
(ind_newscrawl_2011_1M:699503)

1 See Section 4 for details on the source of the examples. Examples are glossed following the Leipzig Glossing Rules. Abbreviations used in the glosses: 1 first person; 2 second person; ABBREV abbreviation; AV active voice prefix; BE copula; CAUS causative; DEM demonstrative; EXCL exclusive; FUT future marker; INCL inclusive; LOC locative; NAME proper name; PASS passive voice prefix; PL plural; POSS possessive; PROG progressive marker; REL relativiser; SG singular.

***Corresponding author: I Made Rajeg**, Universitas Udayana, Kampus Bukit, Jl. Raya Kampus Unud Jimbaran, 80361, Bukit Jimbaran, Bali, Indonesia, E-mail: made_rajeg@unud.ac.id. https://orcid.org/0000-0001-8989-0203
Gede Primahadi Wijaya Rajeg, Universitas Udayana, Kampus Bukit, Jl. Raya Kampus Unud Jimbaran, 80361, Bukit Jimbaran, Bali, Indonesia, E-mail: primahadi_wijaya@unud.ac.id. https://orcid.org/0000-0002-2047-8621
I Wayan Arka, Linguistics Program, School of Culture, History & Language, The Australian National University, Canberra, ACT 0200, Australia; and Universitas Udayana, Kampus Bukit, Jl. Raya Kampus Unud Jimbaran, 80361, Bukit Jimbaran, Bali, Indonesia, E-mail: wayan.arka@anu.edu.au. https://orcid.org/0000-0002-2819-6186

(2) *akhirnya* *laga* *eksebisi* *tersebut* ***di-maju-kan***.
finally match exhibition DEM PASS-move.forward-CAUS
'… finally, the exhibition match is *moved forward* (*to an earlier time*).'
(ind_newscrawl_2016_1M:591217)

The examples in (1) and (2) describe the causation of an event earlier than scheduled. In (1), the Actor (*Dindik Surabaya*) is the subject (SUBJ) of *meN*-prefixed AV *memajukan* while the Undergoer (*jadwal*) is the direct object (OBJ). In (2), the Undergoer is the SUBJ of *di*-prefixed PASS *dimajukan* (with an unexpressed Actor).

This paper investigates the meaning-preserving hypothesis in voice alternation from a quantitative usage-based perspective in the context of two strands of research. The first follows research traditions on voice alternation in Austronesian (AN) languages, including Indonesian, from the perspective of quantitative discourse (Sections 2 and 3). Second, the growing interest within usage-based construction grammar (CxG; Goldberg 2006) in the association between (non-)metaphoric senses and grammatical constructions (Sullivan 2013), such as word class, inflection, syntactic patterning (Deignan 2006), and aspectual prefixation (Sokolova 2013). We investigate Indonesian CAUSED FORWARD/BACKWARD MOTION verbs and the distribution of their (non-) metaphoric senses in AV and PASS (Sections 3 and 4). Section 4 elucidates our approach to combining corpus and experimental data as a methodological paradigm in cognitive linguistics and CxG (Gries 2013). After presenting the quantitative analyses in Section 5, Section 6 discusses the implications for (a) the frequency effect and sense-sensitivity of the meaning-preserving hypothesis in voice alternation (Section 6.1); and (b) the item-specific representation of linguistic knowledge in usage-based CxG (Section 6.2).

2 Overview of research on voice in Austronesian languages

Typological research on AN languages has demonstrated that AN possess a richer voice opposition than simply AV versus PASS as in English (Arka and Ross 2005; Himmelmann 2005). Besides the morphological voice alternation seen in (1) and (2), Indonesian exhibits so-called Undergoer voice (UV), exemplified in (3), where the verb is unmarked morphologically.

(3) *maka* *perayaan-nya* *kita* ***maju-kan*** *satu* *hari.*
therefore celebration-DEM 1PL.INCL move.forward-CAUS one day
'therefore the celebration, we *moved* it *forward* one day.'
(ind_newscrawl_2016_1M:787910)

UV is like PASS, in that the Undergoer (*perayaan* 'celebration') is the SUBJ, but, unlike in PASS, the Actor (*kita* 'inclusive we') is still a core argument. UV is syntactically transitive and is an example of a symmetrical voice system (Arka 2003; Foley 1998; Himmelmann 2005; Riesberg 2014). Given the very low frequency of UV in our corpus compared to AV *meN-* and PASS *di-*, we excluded UV from analysis, but acknowledge its significance in AN voice systems. Note that the unmarked verb can also occur in AV (with the Actor as SUBJ), as seen in (4), especially in "spoken colloquial Indonesian" (Riesberg 2014: 12).

(4) *Kami* *hanya* ***ajukan*** *tiga* *karena* *satu* *saksi* *adalah* *terdakwa.*
1PL.EXCL only propose three because one witness BE defendant
'We only *proposed* (i.e. *nominated*) three [witnesses] because one witness is a defendant.' (the formal, prefixed version is *mengajukan* 'to propose/put forward')
(ind_newscrawl_2016_1M:606293)

Another central issue is AN's typological alignments (e.g. ergative or nominative), which have been subject to corpus-based studies within Hopper and Thompson's (1980) framework. For Indonesian, Wouk (1996) analysed 88 clauses from a narrative corpus of Spoken Jakarta Indonesian to investigate voice selections (i.e. Patient-trigger and Actor-trigger) and test the discourse-ergative status of Spoken Jakarta Indonesian. Wouk (1996) shows that Patient-trigger is significantly associated with higher transitivity than Active-trigger, and also with greater individuation of OBJ.

3 Aims

This paper takes a different spin on Wouk's (1996) finding that AV and PASS have different distributions in discourse. We investigate the association between (non-)metaphoric senses of verbs and their expression in voice morphology. This aim relates to McDonnell's (2016) collostructional analysis (Gries and Stefanowitsch 2004) of the AN language Besemah,[2] where a verbal root demonstrates a statistical preference for a particular voice (see Gries and Stefanowitsch [2004] for similar findings in English). McDonnell (2016: 250) hypothesizes that the semantic properties of the root may condition voice selection. Given the meaning-preserving hypothesis, the null hypothesis would be that all potential senses of a verbal base should be equal in terms of the frequency with which they are expressed in AV and PASS. The alternative hypothesis could build on McDonnell's assumption that a given sense (i.e. semantic property) of a verb would be (dis)preferred in a certain voice. Categorically, the meaning-preserving hypothesis would predict that all senses of a verb can alternate, but it would not predict the preference and entrenchment of a particular sense in a given voice, which is central to the usage-based model of language (Diessel and Hilpert 2016; Langacker 1988).

4 Data and methodology

This study combined corpus data with sentence-production experimental data (Dąbrowska 2009; Newman and Sorenson Duncan 2019). The corpus (53,734,159 tokens; see Table 1) comes from the Indonesian corpus of the Leipzig Corpora Collection (Goldhahn et al. 2012) and consists of shuffled sentences from online news.

We studied transitive CAUSED FORWARD and BACKWARD MOTION verbs based on the following intransitive bases: *aju*,[3] *maju*, *undur*, and *mundur*. *Aju* appears to be a reflex of an old AN form *-atu 'forward, onward; towards the hearer' (Blust and Trussel 2010)[4] and retains its derivation as *maju* in present-day Indonesian[5] with the reflex of the archaic AN middle/stative prefix *m(a)-.[6] The base *undur* 'to move backward' is listed as an intransitive verb in *Kamus Besar Bahasa Indonesia* (2016; KBBI). It relates to *mundur*, which retains the AN middle prefix *m(a)- and the semantics of the old AN form *undur 'retreat, fall back' (Blust and Trussel 2010).[7] *Undur* can occur in transitive constructions with or without the causative suffix *-kan*, while the other bases need *-kan* to appear in order to evoke meanings related to caused motion. These suffixed bases can take *meN-* for AV, *di-* for PASS, or remain unprefixed for UV constructions (Table 2).

Table 1: The names and sizes of the files that make up the corpus.

	Filename	Size (in words)
1	ind_newscrawl_2016_1M	15,702,910
2	ind_newscrawl_2015_300K	4,909,696
3	ind_newscrawl_2012_1M	16,822,496
4	ind_newscrawl_2011_1M	16,299,057

2 Besemah is a Malayic language spoken in the highland of Northern Sumatra, Indonesia.

3 According to *Kamus Besar Bahasa Indonesia* (2016), the root *aju* is pre-categorial (i.e. a bound base or root; Lieber [2010: 34]), meaning that *aju* cannot occur in verbal syntax without affixation.

4 https://www.trussel2.com/ACD/acd-pl_pwmp.htm. We thank the anonymous reviewer for pointing this out.

5 The anonymous reviewer suggests that *maju* 'to walk/move forward' is not a root, but a derived middle verb. While this analysis is appreciated, speakers in present-day Indonesian are unaware of this connection; even KBBI lists *maju* as a root in a different entry from *aju*.

6 The cognates of this old AN middle *m(a)- are also found in other AN languages, such as Balinese (Shibatani and Artawa 2007) and Taiwanese. We thank the anonymous reviewer for this insight.

7 https://www.trussel2.com/ACD/acd-s_u1.htm#5812.

Table 2: Number of occurrences of the bases in their specific voice morphologies in the corpus.

Base	Unprefixed	AV: *meN-*	PASS: *di-*	Total
majukan	57	1,725	151	1,933
mundurkan	8	53	33	94
aju	19	10	0	29
ajukan	763	7,742	4,826	13,331
undur	56	45	384	485
undurkan	5	1,922	12	1,939
Total	908	11,497	5,406	17,811

Table 2 reveals that a given base occurs more frequently in certain morphological patterns than others. We analysed a random sample of 100 concordances for each combination of base and prefix or no prefix, giving a maximum of 300 examples for each base. Wouk (1996: 371) excluded the unprefixed bases since their AV and PASS statuses are difficult to determine morphologically. Our study also found unclear cases of unprefixed bases, but their UV versus AV occurrences can be inferred from word order when the Actor and Undergoer are overt: UV is identifiable when the Undergoer precedes the Actor, which in turn precedes the unprefixed verb, as seen in (3).

The experimental data was gathered using an online questionnaire distributed to 231 undergraduate students in the English departments at Udayana University and Mahasaraswati University in Indonesia, in exchange for course credit. All participants speak Indonesian. The participants were prompted to produce a sentence using the target verbs interspersed randomly within other filler words (total 34 words, including the target verbs); we asked participants for sentences containing the base (e.g. *majukan*), AV (e.g. *memajukan*), and PASS (e.g. *dimajukan*) forms. Participants did not know our target verbs nor the aim of the questionnaire beyond the need to produce sentences. Each word had its own page, and participants pressed "Next" to move to the next word. The questionnaire was closed on the third day after being distributed. We received responses from 118 participants (74.58% female and 24.58% male; one provided no gender), allowing us to retrieve at least 100 random samples for each target verb before discarding irrelevant responses (e.g. if participants responded with a different verb).

To analyse the (non-)metaphoric senses of the verbs, we adopted insights from Conceptual Metaphor Theory (Lakoff and Johnson 1980), viewing metaphor as understanding an abstract domain in terms of a concrete domain. For instance, an abstract domain PLEASURE can be construed as a DECEIVER as in ***tergiur*** *oleh kesenangan* '*lured* by pleasure' and *kesenangan yang* ***menipu*** '*deceiving* pleasure' (Rajeg 2019: 189–191). We coded usages of the verbs as non-metaphoric if their truth condition referred to the physical motion of a concrete entity; otherwise, they were coded as metaphoric (Stefanowitsch and Goschler 2009: 170–171).

(5) *Paman sedang* ***me-mundur-kan*** *mobil=nya untuk parkir*
uncle PROG AV-move.backward-CAUS car=3SG.POSS in.order.to park
'Uncle is *moving* his car *backward* in order to park [it].'
(Elicitation: 99–10.35.31)

(6) *kebijakan dalam* ***me-maju-kan*** *pendidikan di bidang hukum*
policy inside AV-move.forward-CAUS education LOC field law
'… policy in *advancing* law education'
(ind_newscrawl_2016_1M:767269)

Sentence (5) describes a car's actual change of location; the car must have moved from its origin so that (5) can be true. Sentence (6), however, describes a non-actual motion: advancing the state of the OBJ *pendidikan*. For (6) to be true, the OBJ does not have to change location. Thus, (5) was literal while (6) was metaphoric: sentence (6) is metaphoric since *pendidikan* 'education' cannot undergo physical motion, given the basic sense of *memajukan*, whereas (5) indicates a physical change in the car's location. These examples illustrate that

(non-)metaphoric verb usage can be inferred from its argument's referent, but by no means do they indicate that metaphoric meaning is always clear-cut.

We then compared and tested the frequencies of (non-)metaphoric senses in AV and PASS (Levshina 2015: Ch. 9) for statistical significance using a chi-square test, or Fisher's exact test when the distributional assumptions for a chi-square test were not met (Levshina 2015: 213–214). For the data and R scripts for the quantitative analyses, see the supplementary materials in Rajeg et al. (2021).

5 Results

5.1 Analysis for *majukan*, *memajukan*, and *dimajukan*

In the corpus data, there are 48 tokens of unprefixed *majukan* (after removing nine duplicates and typos): 8.33% ($N = 4$) occur in UV, and the rest are the unprefixed AV, which we excluded following Wouk (1996).

Figure 1 shows that UV is the least frequent voice in which *majukan* is found, compared to AV and PASS. No literal sense was found in the sample for AV *memajukan*, however one token was identified in the sample for PASS *dimajukan*; for this reason, we excluded the literal sense in the corpus analysis. Figure 2 visualizes the distribution of senses for *memajukan* and *dimajukan*.

Figure 2 demonstrates a highly significant distribution and robust effect size (see the Cramér's *V* values[8]). The 'advancing' sense, as seen in (6), occurs more frequently in AV than PASS. The temporal 'cause to happen earlier' sense, exemplified in (2) and (1), is more predominant in PASS than AV. The 'proposing' sense, shown in (7), is also higher in PASS.

(7) *figur* *muda* *yang* *akan* ***di-maju-kan*** *pada* *Pilpres* *2009*
figure young REL FUT PASS-move.forward-CAUS at presidential.election 2009
'… young figure who will be *proposed/nominated* (i.e. *put forward*) for the 2009 presidential election' (ind_newscrawl_2011_1M:939570)

The association plot (Figure 3) shows these asymmetries more intuitively. The bluish bars rising above the line indicate a positive deviation from chance distribution, while the reddish bars falling below the line indicate a negative deviation. The luminosity signals association strength: the darker, the stronger (Levshina 2015: 220). Thus, 'advancing' strongly prefers AV; 'cause to happen earlier' and 'proposing' strongly prefer PASS.

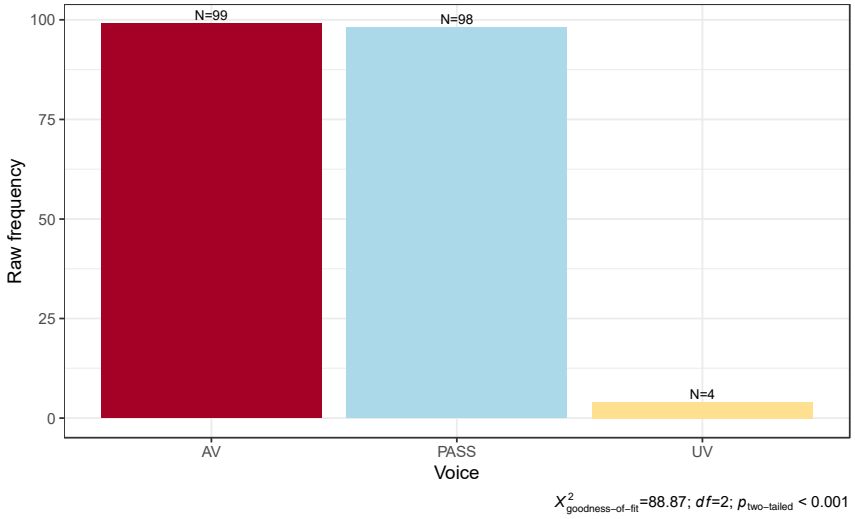

Figure 1: Distribution of *majukan* across voice (corpus data).

8 Cramér's *V* values are interpreted following Levshina (2015: 209): $0.1 \leq V < 0.3$ indicates a small effect; $0.3 \leq V < 0.5$ indicates a moderate effect; $V > 0.5$ indicates a large or strong effect.

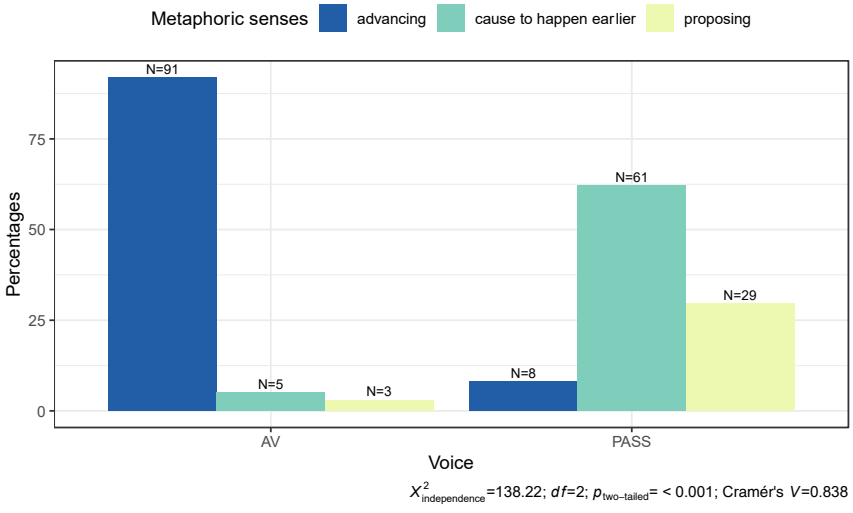

Figure 2: Senses of *majukan* across voice (corpus data).

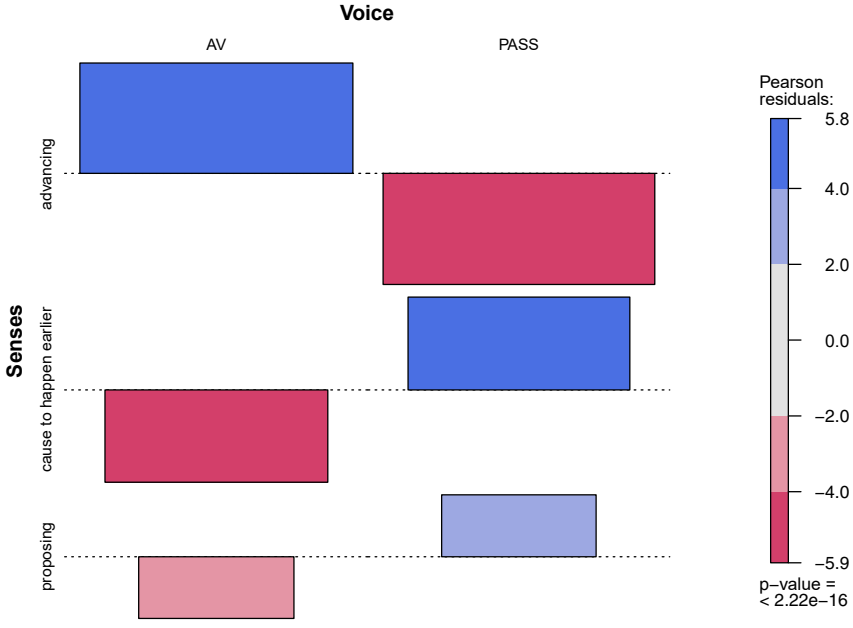

Figure 3: Association plot for *majukan* (corpus data).

Figure 4 displays the sentence-production data and reveals converging results, especially for 'advancing' and temporal senses. Participants did not produce sentences expressing 'proposing', parallel to (7). As in the corpus study, we compared the prefixed AV and PASS verbs, but included production data with literal senses. 'Advancing' and temporal senses are strongly associated with AV and PASS respectively, as in the corpus data: 'advancing' occurs significantly more often in AV but is dispreferred in PASS over 'cause to happen earlier'. The literal sense of 'caused forward motion' is more frequent in PASS. The magnitude of these effects is strong (Cramér's V = 0.872).

Figure 5 accentuates this correlation, mirroring the corpus results for 'advancing' and temporal senses (Figure 3). Participants appear to store detailed semantic preferences for *majukan* in AV and PASS.

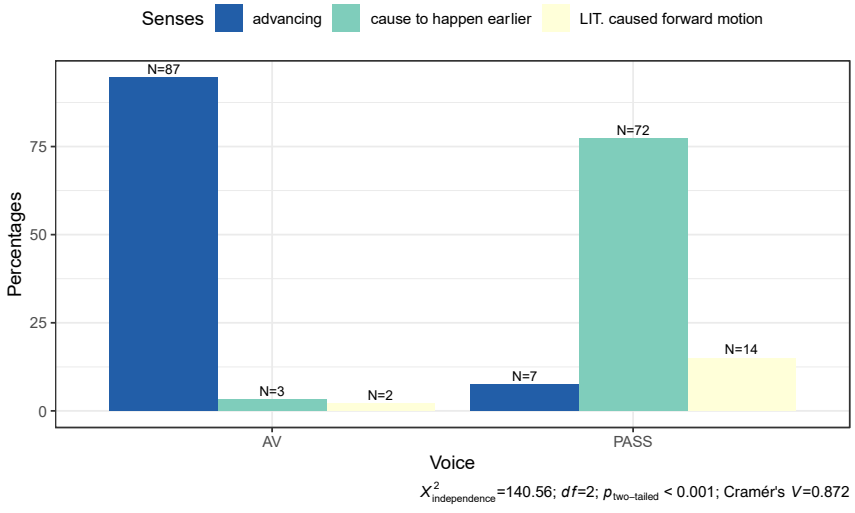

Figure 4: Senses of *majukan* across voice (sentence-production data).

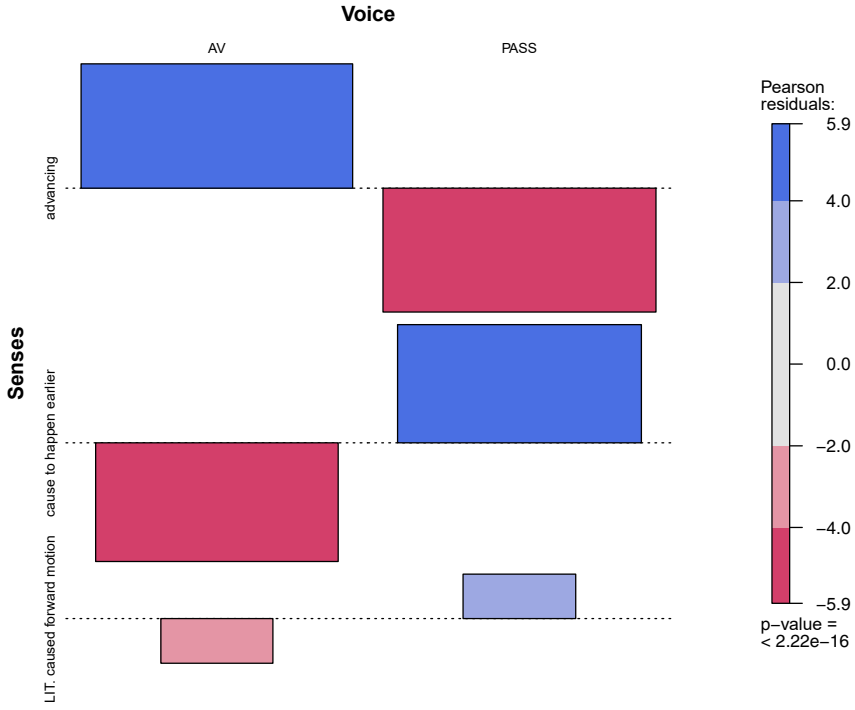

Figure 5: Association plot for *majukan* (sentence-production data).

5.2 Analysis for *mundurkan*, *memundurkan*, and *dimundurkan*

In the corpus data, the unprefixed *mundurkan* 'cause to move something backward' occurs in seven tokens (excluding duplicates) with only one in UV and the remaining in AV; they are excluded from the analysis. Figure 6 visualizes the distribution of senses for AV *memundurkan* and PASS *dimundurkan*.

The 'postpone' sense, as seen in (8), has nearly equal distribution in AV (64.7%; $N = 33$) and PASS (66.7%; $N = 22$). The literal sense ($N = 16$), exemplified in (5), appears more frequently in AV, while the sense 'withdraw someone from a (structural) position; cause someone to step down' ($N = 10$), shown in (9), is lower in AV (7.84%; $N = 4$) than in PASS (18.2%; $N = 6$). An additional sense that was attested relates to 'retreat/change one's mind'; see example (10).

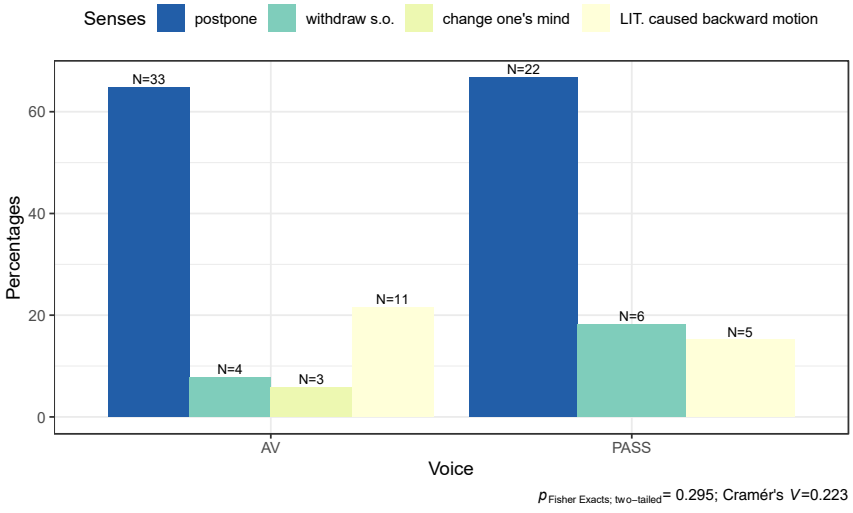

Figure 6: Senses of *mundurkan* across voice (corpus data).

(8) *Seharusnya* *pertemuan* *dengan* *SFC* *tanggal* *22 Maret,*
supposedly meeting with NAME date 22 March
tetapi ***di-mundur-kan*** *menjadi* *29 Maret.*
but PASS-move.backward-CAUS become 29 March
'The meeting with SFC is supposedly on 22 March, but [it has been] *moved backward* to 29 March' (ind_newscrawl_2011_1M: 892106)

(9) *menurutnya,* *Andi* *harus* *segera* *mundur* *atau* ***di-mundur-kan.***
in.h(is/er).opinion NAME must soon move.backward or PASS-move.backward-CAUS
'According to him/her, Andi must soon *step back/resign* or [someone] will *cause* [him] *to resign*' (ind_newscrawl_2012_1M: 322769)

(10) *sampai* ***me-mundur-kan*** *keberanian=nya* *dalam* *maju* *berjuang.*
until AV-move.backward-CAUS bravery=3SG.POSS inside move.forward fight
'[Let alone] until *moving back/retreating/changing* his/her bravery in going forward fighting' (ind_newscrawl_2016_1M: 41051)

The distributional asymmetry of these senses across voice is small and not statistically significant. This is depicted visually in Figure 7 via the grey shading, and highlights the meaning-preserving trait of *mundurkan* across voice types.

By way of contrast, the sentence-production data (Figure 8) shows a highly significant distribution with only a moderate effect. What stands out from Figure 8 is that the literal sense is more frequent in AV than in PASS, whereas 'postpone' is more frequent in PASS than in AV. These asymmetries represent strong deviations from the expected frequencies (Figure 9). This suggests that participants more strongly represent the literal sense in AV and 'postpone' in PASS, while the genre of the corpus data may have augmented the prominence of 'postpone' over the literal sense in AV (Figure 6). The sentence-production data reveals another sense that is absent in the corpus, namely 'deteriorate' (e.g. *memundurkan kualitas/kesejahteraan/perekonomian* '*move back/deteriorate* the quality/prosperity/economy'). So far, the results for *mundurkan* (this section) and *majukan* (Section 5.1) suggest the sense-sensitivity of voice alternation, where each sense is associated with a different voice (see Bernolet and Colleman [2016] for similar evidence in Dutch dative alternation).

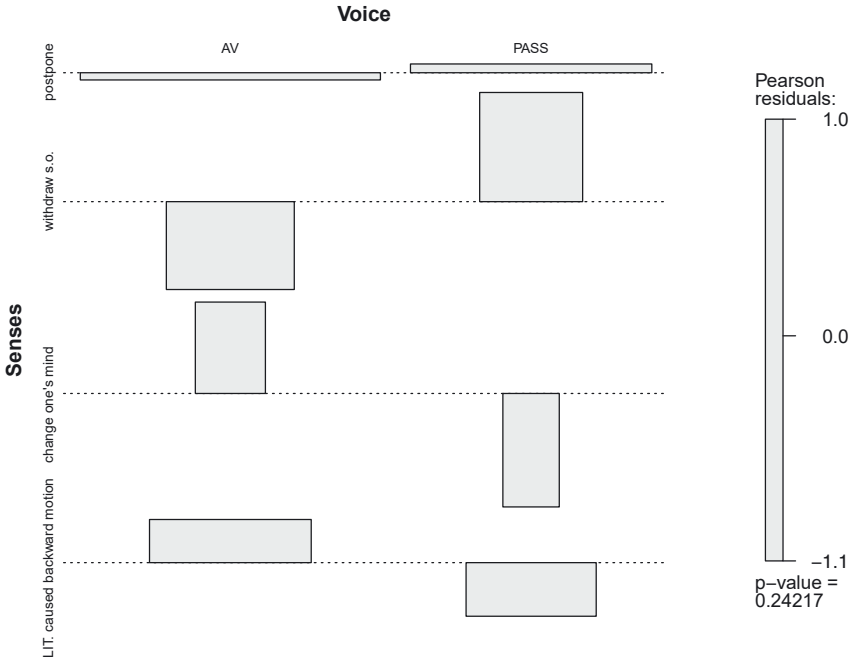

Figure 7: Association plot for *mundurkan* (corpus data).

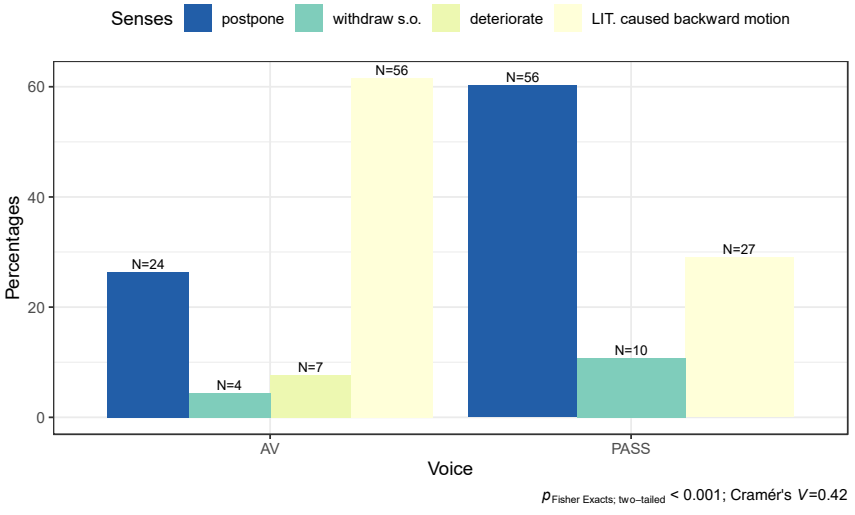

Figure 8: Senses of *mundurkan* across voice (sentence-production data).

5.3 Analysis for *ajukan*, *mengajukan*, and *diajukan*

All nine tokens for the root *aju* found in the corpus data function as modifiers meaning 'advanced' (e.g. *tim aju* '*advanced* team'). The AV *mengaju* is also attested, but we are treating those examples as mistakes, as the appropriate words in the examples where it is used would be *mengaku* 'to admit something' and *mengacu* 'to refer to'. The PASS *diaju* is absent in the corpus. These AV and PASS forms are also absent from the KBBI entry.

The verbal form for *aju* is conventionally suffixed with *-kan,* and this base *ajukan* is attested in 763 tokens in the corpus; the number of UV tokens is significantly lower than the number of tokens of AV *meN-* and PASS *di-* (Figure 10).

The most frequent use of *ajukan* in KBBI is metaphoric, 'to propose/put forward an idea/argument/refutation/candidate', while the second sense is physical, 'to bring to the front'. All tokens for *ajukan* in AV and PASS in the corpus convey 'proposing', as seen in (4) ($X^2_{\text{goodness-of-fit}} = 0.005$; $df = 1$; $p_{\text{two-tailed}} = 0.943$). The same is true for the sentence-production data ($X^2_{\text{goodness-of-fit}} = 0.0497$; $df = 1$; $p_{\text{two-tailed}} = 0.824$). This indicates that *ajukan* is meaning preserving across voice types.

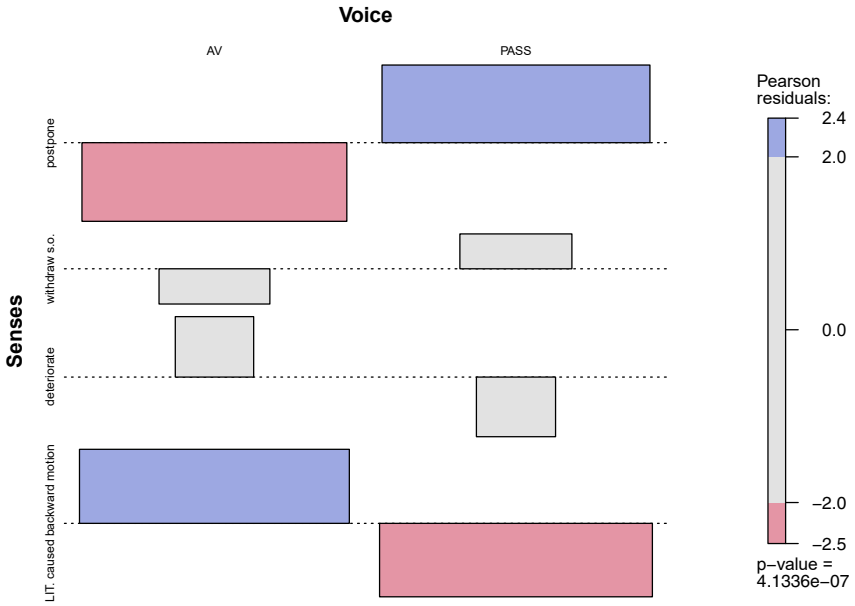

Figure 9: Association plot for *mundurkan* (sentence-production data).

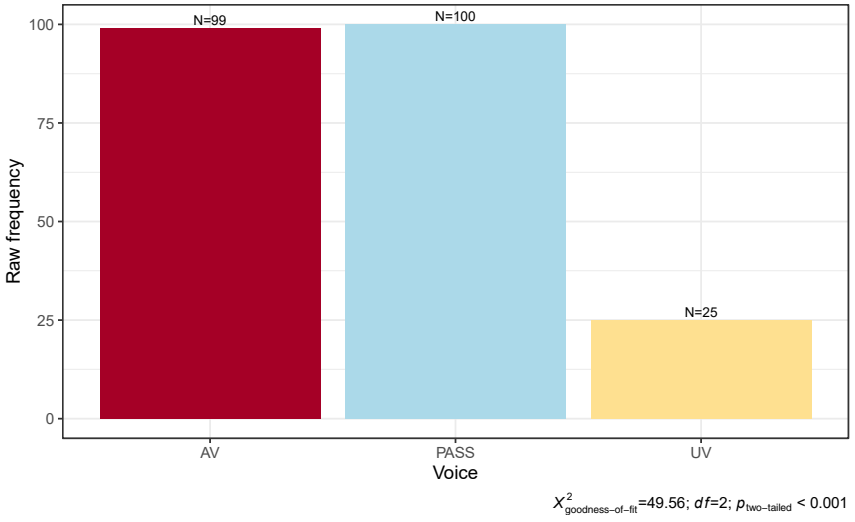

Figure 10: Distribution of *ajukan* across voice (corpus data).

5.4 Analysis for *undur*, *undurkan*, *mengundurkan*, and *diundurkan*

Figure 11 shows the distributional asymmetry of the bases *undur* and *undurkan* across voice types. The AV *mengundurkan* is much more frequent than the AV *mengundur*. In contrast, the PASS *diundur* is much more frequent than the PASS *diundurkan*. The UV form, only attested with *undur*, is significantly lower than expected compared to AV and PASS with *undur* ($X^2_{\text{goodness-of-fit}}$ = 121.29; *df* = 2; $p_{\text{two-tailed}}$ < 0.001).

5.4.1 Undur

The AV *mengundur* (*N* = 24) and PASS *diundur* (*N* = 99) both express a meaning of 'postpone'. *Undur* in AV and PASS is thus categorically meaning preserving, though the 'postpone' sense is significantly higher in PASS than AV ($X^2_{\text{goodness-of-fit}}$ = 45.73; *df* = 1; $p_{\text{two-tailed}}$ < 0.001) due to a higher token count for PASS than AV (Figure 11). Similarly, all transitive tokens of AV *mengundur* (*N* = 74) and PASS *diundur* (*N* = 91) in the sentence-production data convey 'postpone' without significant differences in voice type ($X^2_{\text{goodness-of-fit}}$ = 1.75; *df* = 1; $p_{\text{two-tailed}}$ = 0.186) and hence are meaning preserving.

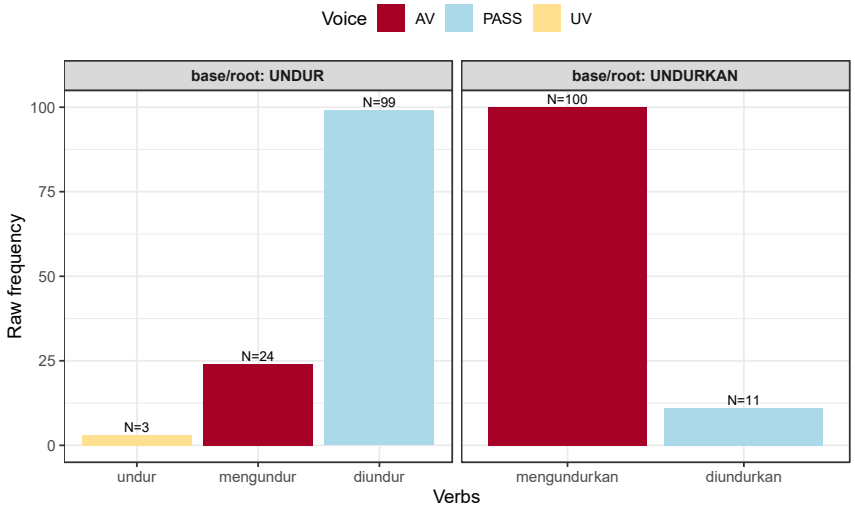

Figure 11: Distribution of *undur* and *undurkan* across voice (corpus data).

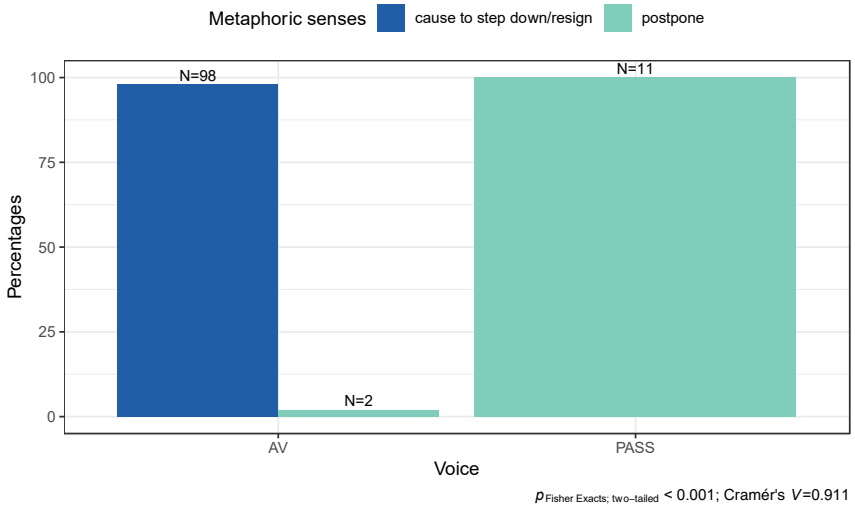

Figure 12: Senses of *undurkan* across voice (corpus data).

5.4.2 Undurkan

Unlike *undur*, the suffixed form *undurkan* predominantly expresses 'cause to step down/resign' (88.29%; $N = 98$) rather than 'postpone' (11.71%; $N = 13$). No literal sense is attested in the corpus. Figure 12 shows the distributional asymmetry of these senses across voice types.

The strong effects include the preference for 'postpone' in PASS and its dissociation with AV, and the dispreference for 'step down' in PASS (Figure 13). The high frequency of 'step down' in AV is due to the fixed verb phrase *mengundurkan diri* 'to cause oneself to step back; resign'; its PASS equivalent is unattested.

The uneven distribution of senses across voice in the sentence-elicitation data (Figure 14) is also significant. The association plot (Figure 15) demonstrates that 'postpone' is strongly preferred in PASS while 'cause to step down/resign' is preferred in AV (again due to the fixed phrase *mengundurkan diri*). As in the corpus sample, no literal, physical meaning is produced by the participants for *undurkan*.

The analyses of *majukan*, *mundurkan*, and *undur*(*kan*) in AV and PASS reveal some parallelism in their senses. The opposing temporal senses (i.e. 'happen earlier' vs. 'postpone') demonstrate an affinity to PASS voice types in nearly all verbs studied (Table 3).

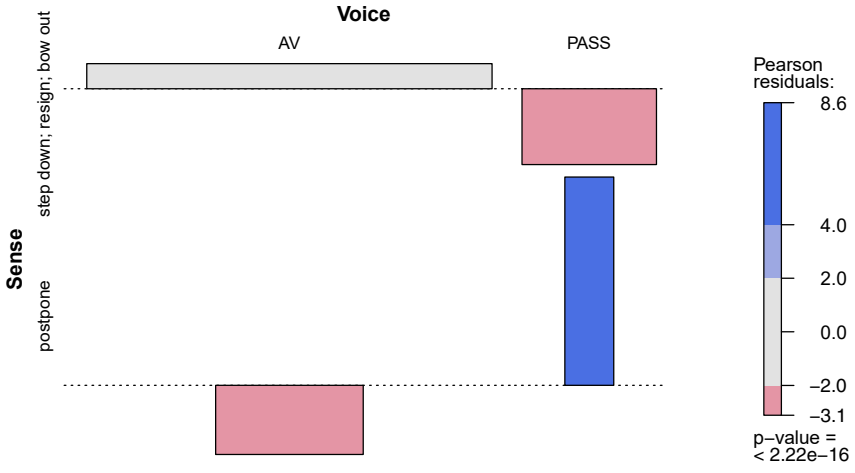

Figure 13: Association plot for *undurkan* (corpus data).

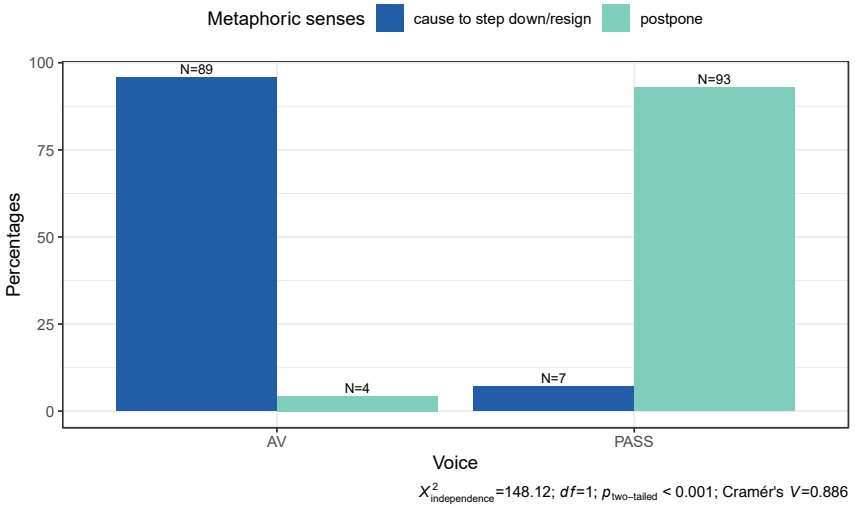

Figure 14: Senses of *undurkan* across voice (sentence-production data).

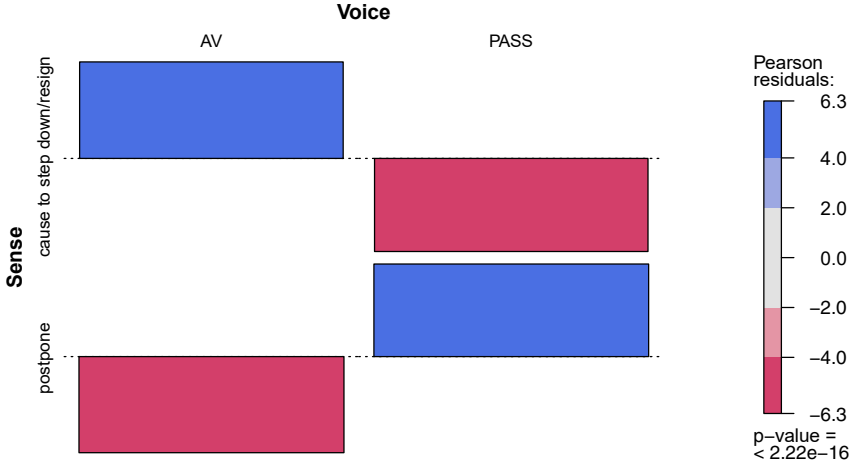

Figure 15: Association plot for *undurkan* (sentence-production data).

Table 3: Parallelism across words for the interconnection of PASS and temporal meanings. "PASS > AV" indicates that the observed frequency of that sense is greater than expected in PASS compared to AV.

Temporal sense	Base	Data	Observed versus expected	Significance (and effect size)
'happen earlier'	*majukan*	corpus	PASS > AV	$p < 0.001$; Cramér's $V = 0.838$
		elicitation	PASS > AV	$p < 0.001$; Cramér's $V = 0.872$
'postpone'	*mundurkan*	corpus	PASS > AV	$p > 0.1$, ns; Cramér's $V = 0.223$
		elicitation	PASS > AV	$p < 0.001$; Cramér's $V = 0.42$
	undur	corpus	PASS > AV	Goodness-of-fit: $p < 0.001$
		elicitation	PASS > AV	Goodness-of-fit: $p > 0.1$, ns
	undurkan	corpus	PASS > AV	$p < 0.001$; Cramér's $V = 0.911$
		elicitation	PASS > AV	$p < 0.001$; Cramér's $V = 0.886$

6 Discussion

6.1 Meaning-preserving hypothesis and voice alternation

This study highlights the dynamics of the meaning-preserving hypothesis for voice alternation in Indonesian. Quantitative analyses revealed a statistical tendency for a given sense to be associated with a particular voice morphology. This tendency indicates that a verb's propensity to convey a particular sense involves distinct (morphological) constructions (Deignan 2006; Sokolova 2013), and is not only due to the semantics of its nominal collocates.

We have demonstrated that certain senses exhibit strong associations with certain voice forms, both in the corpus data and in the sentence-production data (e.g. a relatively strong correlation between temporal meanings and PASS *di-* compared to AV *meN-*; see Table 3). These findings provide initial support for McDonnell's (2016) proposal (Section 3) that the statistical verb bias for a voice type could be driven by the verb's semantic property. Our findings further support previous studies that have shown how syntactic alternations, such as the Dutch dative alternation and the Russian locative alternation, are sensitive to both verb senses (Bernolet and Colleman 2016), and the interaction between metaphor and constructions (Sokolova 2013). We have revealed that sense-sensitivity exists in Indonesian morphological voice alternation and argued that such alternations should be relativized to verb senses. Our results converge with recent work on the Indonesian HITTING verb *kena* 'be hit'; one of the senses of this verb is directly constructed in PASS without corpus evidence for the same sense in AV (Rajeg et al. 2020). This result corroborates the idea that PASS can have distinct semantic constraints from AV (Hilpert 2014: 41). More broadly, our study offers a new perspective into voice alternation in AN.

6.2 Usage-based construction grammar: implication and explanation

Our findings contribute to, and can be naturally explained within, the tenets of usage-based CxG (Diessel 2015; Goldberg 2006; Hilpert 2014). The first tenet is that frequency effects may impact the cognitive representation of linguistic knowledge (Diessel 2016; Diessel and Hilpert 2016; Hilpert and Diessel 2016). Frequent co-occurrence of a particular form with a particular meaning drives the automatization and conventionalization of that form-meaning pairing in production and processing (Diessel and Hilpert 2016). Some converging findings between our experimental and corpus data indicate that participants demonstrate some representation of the strong association between a given voice form of the verbs and the sense predominantly expressed in that form (e.g. frequent co-occurrences of PASS *dimundurkan* 'cause to move back' with time/event nouns to express 'postpone'). This reflects the view that participants' knowledge of the semantic preference of these verbs in a given voice is grounded in their frequent exposure to concrete usages of these verbs (Dąbrowska 2009). Such high token co-occurrence frequencies of the verbs in certain contexts

lead to their "automatization" (Diessel 2016: 226), building associative links between the verbs' voice form and the sense expressed. Statistical tendencies, as employed in this paper, provides one measure for the strength of the associative links (Diessel 2016: 226–229).

The second tenet of usage-based CxG is the idea that grammar consists of "a unified repository of form-meaning pairings", that is, constructions (Hilpert and Diessel 2016: 58). These encompass not only abstract grammatical constructions but also concrete tokens of linguistic units (e.g. specific collocations, partially schematic multi-word expressions), including their frequencies (Goldberg 2006: 45). Constructions form hierarchical networks and are connected through different links (e.g. inheritance, polysemy, and subpart links; Hilpert and Diessel 2016). In usage-based CxG, we can investigate usage patterns of specific verbs and not just abstract grammatical constructions (Goldberg 2006: 45).

Converging evidence from our corpus and experimental data suggests that participants retain item-specific knowledge of the verbs (Goldberg 2006: 49, 56; Hilpert 2014: 66), particularly the semantic bias of a verb in one voice construction over the other, given that AV and PASS are available for the senses. A previous study on English EAT and DRINK verbs (Newman and Rice 2006) showed that even different inflections of these verbs have different co-occurrence behaviours. Evidence for such item-specific knowledge also comes from studies on child language acquisition (Tomasello 2000, 2003). The conventionality of the expression of a certain sense in a certain voice could not be predicted by the meaning-preserving hypothesis, since the hypothesis implicitly assumes an equal status for the senses to be expressed in all voice types, while specific usage patterns of the verbs reveal distributional asymmetries. The conventional usage of specific verbs must be accounted for as part of speakers' linguistic repertoire and not only the abstract patterns (Goldberg 2006: 56).

The hierarchical network model of linguistic knowledge (Hilpert 2014: 57–67) and the role of frequency effects in usage-based CxG, including construction morphology (CxM; Booij 2010), can handle the fact that voice alternation is more than simply a re-alignment of syntactic function and semantic role. In CxM, morphologically complex words are viewed as morphological constructions (Booij 2010: 17), having holistic morphosyntactic, morphosemantic, and usage properties (Hilpert 2014: 80). Specific AV and PASS prefixed forms (e.g. *memajukan* and *dimundurkan*) are still connected to their constructional schemas, namely *meN*-x-*kan* and *di*-x-*kan* (where "x" is the base). These schemas emerge as generalizations over concrete tokens of complex words with the same patterns (e.g. *memundurkan*, *dimundurkan*, *mengajukan*, *diajukan*; Booij 2010). The predictable properties inherited from these schemas by the specific words are the schematic meaning of the voice prefixes *meN-* and *di-* and the suffix *-kan*. The network model in CxG and CxM allows specific words to have their idiosyncrasy (e.g. semantic preference and frequency) that may not be fully predicted by, but is still connected to, the schema (Hilpert 2014: 80).

The *-kan* suffix, when unified with an intransitive base *maju* 'move forward', contributes its causative meaning, Actor/Causer role, and transitive argument-structure to *majukan* 'to cause to move forward'; the intransitive SUBJ fuses with the OBJ role of *-kan* (Arka 1993). The unification of the prefix *meN-* with transitive *majukan* accentuates that its SUBJ is the Causer/Actor irrespective of the words specifying this slot, while *di-* specifies that the SUBJ is the Undergoer (Arka and Manning 2008; Gil 2002). The strong association between the complex AV verb *memajukan* and 'advancing' or between PASS *dimajukan* and temporal meaning, for instance, is not predicted by only knowing the semantics of those morphological formatives, but also via recurrent collocational patterns of these verbs with certain types of OBJ. The specific verbs also inherit a prominence effect depending on the AV SUBJ and PASS SUBJ. Given the temporal and 'advancing' senses of *majukan*, for instance, it could be that in the former it is the event or time argument that is considered to be more profiled (Maldonado 2007: 833), rather than the Actor advancing the schedule; hence its SUBJ status and the preference for PASS. For the 'advancing' sense, the active nature of such an event may need to be highlighted, hence its preference for AV. Despite such analyses, general semantic or grammatical rules cannot predict the co-occurrence patterns of (non-)metaphoric senses and the voice construction of verbs (Diessel 2016: 213). Why is the temporal sense not preferred in AV *memajukan* over PASS *dimajukan*, or vice versa for 'advancing', given that the time/event and abstract/institution arguments have the same opportunity to be more prominent? Sentence-production data suggests that speakers may have encountered the temporal sense more frequently in PASS than AV, and such frequency effects are attested in the corpus sample (cf. Diessel 2016: 213).

7 Summary

The study of (non-)metaphoric senses of CAUSED FORWARD/BACKWARD MOTION verbs and voice alternation in Indonesian shows that

- (Non-)metaphoric senses of some verbs interact with the voice construction in which the verb occurs (e.g. temporal is strongly associated with PASS). This demonstrates that the meaning-preserving hypothesis is more nuanced and should be probabilistically relativized to verbs senses.
- Corpus findings on voice-sense interactions converge to a large extent with the experimental data, suggesting the strong representation of such item-specific, form-meaning pairings as part of speakers' linguistic knowledge.

More generally, this study touches on a central question in linguistics regarding the relation between form (e.g. voice morphologies of a verb) and meaning (e.g. the senses associated with those forms; Janda 2016).

Acknowledgements: We sincerely thank the anonymous reviewers for their constructive feedback that significantly improved the paper. We are grateful to the English Department students (classes of 2017, 2018, 2019, and 2020) at Udayana University and Mahasaraswati University for participating in the sentence-production experiment. For discussion and comments, we thank Alice Gaby, Poppy Siahaan, Karlina Denistia, Prihantoro, and participants at the UCREL Corpus Research Seminar series. We also thank John Newman for introducing the sentence-production experiment to us, and Charbel El-Khaissi for proofreading the paper. Finally, we are truly grateful for the waiver scheme granted by De Gruyter to publish the paper open access.

References

Arka, I Wayan. 1993. *Morpholexical aspects of the -kan causative in Indonesian*. Sydney, Australia: University of Sydney Masters thesis.

Arka, I Wayan. 2003. Voice systems in the Austronesian languages of Nusantara: Typology, symmetricality, and undergoer orientation. *Linguistik Indonesia* 21(1). 113–139.

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.

Arka, I Wayan & Malcolm Ross. 2005. Introduction. In I Wayan Arka & Malcolm Ross (eds.), *The many faces of Austronesian voice systems: Some new empirical studies*, 1–15. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University.

Bernolet, Sarah & Timothy Colleman. 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.

Blust, Robert & Stephen Trussel. 2010. The Austronesian comparative dictionary: Web edition. Updated 21 June 2020. http://www.trussel2.com/acd/ (accessed 29 November 2021).

Booij, Geert. 2010. *Construction morphology*. Oxford: Oxford University Press.

Dąbrowska, Ewa. 2009. Words as constructions. In Vyvyan Evans & Stephanie Pourcel (eds.), *New directions in cognitive linguistics* (Human cognitive processing 24), 214–237. Amsterdam; Philadelphia: John Benjamins.

Deignan, Alice. 2006. The grammar of linguistic metaphors. In Anatol Stefanowitsch & Stefan Th. Gries (eds.), *Corpus-based approaches to metaphor and metonymy*, 106–122. Berlin: Mouton de Gruyter.

Diessel, Holger. 2015. 14. Usage-based construction grammar. In Ewa Dąbrowska & Dagmar Divjak (eds.), *Handbook of cognitive linguistics*, 296–322. Berlin: De Gruyter Mouton.

Diessel, Holger. 2016. Frequency and lexical specificity in grammar: A critical review. In Heike Behrens & Stefan Pfänder (eds.), *Experience counts: Frequency effects in language*. Berlin: De Gruyter.

Diessel, Holger & Martin Hilpert. 2016. Frequency effects in grammar. In Mark Aronoff (ed.), *Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press.

Foley, William A. 1998. Symmetrical voice systems and precategoriality in Philippine languages. *Paper presented at the workshop on voice and grammatical functions in Austronesian languages, LFG98, Brisbane, Australia, 1 July*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.7.7620&rep=rep1&type=pdf (accessed 22 October 2020).

Gil, David. 2002. The prefixes *di-* and *N-* in Malay/Indonesian dialects. In Fay Wouk & Malcolm Ross (eds.), *The history and typology of western Austronesian voice systems*, 241–283. Canberra: Pacific Linguistics.

Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
Goldhahn, Dirk, Thomas Eckart & Uwe Quasthoff. 2012. Building large monolingual dictionaries at the Leipzig corpora collection: From 100 to 200 languages. In *Proceedings of the 8th Language Resources and Evaluation Conference (LREC) 2012*, Istanbul, 759–765. http://www.lrec-conf.org/proceedings/lrec2012/pdf/327_Paper.pdf (accessed 6 March 2014).
Gries, Stefan Th. 2013. Data in construction grammar. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford handbook of construction grammar*, 93–108. Oxford: Oxford University Press.
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. 2014. *Construction grammar and its application to English*. Edinburgh: Edinburgh University Press.
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: De Gruyter.
Himmelmann, Nikolaus. 2005. The Austronesian languages of Asia and Madagascar: Typological characteristics. In K. Alexander Adelaar & Nikolaus Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar*, 110–181. London: Routledge.
Hopper, Paul J. & Sandra A. Thompson. 1980. Transitivity in grammar and discourse. *Language* 56(2). 251–299.
Janda, Laura A. 2016. Linguistic profiles: A quantitative approach to theoretical questions. *Język i metoda* 3. 127–145.
Kamus Besar Bahasa Indonesia (KBBI). 2016. Badan Pengembangan dan Pembinaan Bahasa, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia. https://kbbi.kemdikbud.go.id/Beranda (accessed 29 November 2021).
Kroeger, Paul. 2004. *Analyzing syntax: A lexical-functional approach*. Cambridge: Cambridge University Press.
Kroeger, Paul. 2005. *Analyzing grammar: An introduction*. Cambridge: Cambridge University Press.
Lakoff, George & Mark Johnson. 1980. *Metaphors we live by*. Chicago: University of Chicago Press.
Langacker, Ronald W. 1988. A usage-based model. In Brygida Rudzka-Ostyn (ed.), *Topics in cognitive linguistics*, 127–161. Amsterdam: John Benjamins.
Levshina, Natalia. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins.
Lieber, Rochelle. 2010. *Introducing morphology*. Cambridge: Cambridge University Press.
Maldonado, Ricardo. 2007. Grammatical voice in cognitive grammar. In Dirk Geeraerts & Cuyckens Hubert (eds.), *The Oxford handbook of cognitive linguistics*, 829–868. Oxford: Oxford University Press.
McDonnell, Bradley. 2016. *Symmetrical voice constructions in Besemah: A usage-based approach*. Santa Barbara: University of California, Santa Barbara PhD dissertation.
Newman, John & Sally Rice. 2006. Transitivity schemas of English EAT and DRINK in the BNC. In Stefan Th. Gries & Anatol Stefanowitsch (eds.), *Corpora in cognitive linguistics: Corpus-based approaches to syntax and lexis* (Trends in linguistics: Studies and monographs 172), 225–260. Berlin: Mouton de Gruyter.
Newman, John & Tamara Sorenson Duncan. 2019. The subject of ROAR in the mind and in the corpus: What divergent results can teach us. *Linguistica Atlantica* 37(1). 1–27.
Rajeg, Gede Primahadi Wijaya. 2019. *Metaphorical profiles and near-synonyms: A corpus-based study of Indonesian words for Happiness*. Clayton, VIC: Monash University PhD thesis.
Rajeg, Gede Primahadi Wijaya, I Made Rajeg & I Wayan Arka. 2020. Corpus-based approach meets LFG: The puzzling case of voice alternations of *kena*-verbs in Indonesian. In Miriam Butt & Toivonen Ida (eds.), *Proceedings of the LFG'20 conference, on-line*, 307–327. Stanford, CA: CSLI. https://web.stanford.edu/group/cslipublications/cslipublications/LFG/LFG-2020/lfg2020-rra.pdf (accessed 29 November 2021).
Rajeg, I Made, Gede Primahadi Wijaya Rajeg & I Wayan Arka. 2021. Supplementary materials for "Corpus linguistic and experimental studies on the meaning-preserving hypothesis in Indonesian voice alternations". *Open Science Framework*. https://doi.org/10.17605/OSF.IO/QF38H.
Riesberg, Sonja. 2014. *Symmetrical voice and linking in western Austronesian languages*. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9781614518716.
Shibatani, Masayoshi & Ketut Artawa. 2007. The middle voice in Balinese. In Shoichi Iwasaki, Andrew Simpson, Karen Adams & Sidwell Paul (eds.), *SEALSXIII: Papers from the 13th meeting of the Southeast Asian linguistics Society*, 239–261. Canberra: Pacific Linguistics.
Sokolova, Svetlana. 2013. Verbal prefixation and metaphor: How does metaphor interact with constructions? *Journal of Slavic Linguistics* 21(1). 171–204.
Stefanowitsch, Anatol & Juliana Goschler. 2009. Sex differences in the usage of spatial metaphors: A case study of political language. In Kathleen Ahrens (ed.), *Politics, gender and conceptual metaphors*, 166–183. Chippenham: Palgrave Macmillan.
Sullivan, Karen. 2013. *Frames and constructions in metaphoric language*. Amsterdam: John Benjamins.
Tomasello, Michael. 2000. First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics* 11(1–2). 61–82.
Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
Wouk, Fay. 1996. Voice in Indonesian discourse and its implications for theories of the development of ergativity. *Studies in Language* 20(2). 361–410.