



PROSIDING

**Seminar Nasional Bahasa Ibu (SNBI XII)
Lokakarya Pelestarian Bahasa Ibu (LPBI II)**

Fakultas Ilmu Budaya Universitas Udayana
Denpasar, 6-7 Februari 2020

PROSIDING

SEMINAR NASIONAL BAHASA IBU XII

DAN

LOKAKARYA PELESTARIAN BAHASA IBU II

" DOKUMENTASI DAN REVITALISASI BAHASA LOKAL SEBAGAI IDENTITAS LOKAL DI DUNIA GLOBAL "

Penyunting

I Wayan Mulyawan, S.S., M.Hum.

Dr. Made Sri Satyawati, S. S., M. Hum.

Prof. Dr. Drs. I Nyoman Suparwa, M.Hum.

Prof. Dr. Ketut Artawa, M.A.

Dr. Dra. Maria Matildis Banda, M.S.

**PROSIDING
SEMINAR NASIONAL BAHASA IBU XII
DAN LOKAKARYA PELESTARIAN BAHASA IBU II**

"DOKUMENTASI DAN REVITALISASI BAHASA LOKAL SEBAGAI IDENTITAS LOKAL DI DUNIA GLOBAL "

Penyunting:

I Wayan Mulyawan, S.S., M.Hum.

Dr. Made Sri Satyawati, S. S., M. Hum.

Prof. Dr. Drs. I Nyoman Suparwa, M.Hum.

Prof. Dr. Ketut Artawa, M.A.

Dr. Dra. Maria Matildis Banda, M.S.

ISBN:

978-602-294-403-4

Sampul

I Nyoman Angga Wirama

Tata Letak dan Typeset:

I Gusti Ngurah Parthama, S.S., M.Hum.

Gede Primahadi Wijaya Rajeg, Ph.D.

Kadek Putri Yamayanti, S.S.

Penerbit:

Udayana University Press

Percetakan:

89printing.com

Jl. Kerthapetasikan IV No. 1 Sidakarya - Denpasar – Bali

Hak Cipta dilindungi Undang-Undang

Dilarang mengutip atau memperbanyak sebagian atau seluruh isi buku ini tanpa izin
tertulis dari penerbit

DAFTAR ISI

KATA PENGANTAR	i
DAFTAR ISI.....	ii
Tipologi Tataurutan Kata Bahasa Minangkabau: Perlukah Dokumentasi dan Revitalisasi Bahasa?	1
Jufrizal	1
Pengenalan Istilah Budaya Lokal Melalui Penerjemahan.....	19
Ida Ayu Made Puspani.....	19
Pemanfaatan Video pada Pengajaran Kosakata Bahasa Bali: Studi Kasus Siswa Kelas 3 SD Tunas Daud.....	24
Ketut Santi Indriani	24
Makna Metafora pada Judul Berita Media Cetak.....	31
I Gusti Ngurah Parthama; Ni Ketut Alit Ida Setianingsih	31
Bahasa dan Aksara Daerah di Ranah Publik Pasca-UU RI No. 24 Th 2009: Kajian Linguistik Forensik	36
I Wayan Pastika.....	36
Ideologi Penggunaan Bahasa Bali di Kelurahan Ubud.....	44
Ni Luh Sutjiati Beratha	44
Pemuliaan Sawah dalam “Kuttara Kanda Dewa Purana Bangsul”	61
I Ketut Eriadi Ariana; I Made Sukma Manggala; I Dewa Gde Alit Satria Wibhawa	61
Bahasa Bali: Kaitannya dengan Maksim Penerimaan dan Maksim Kerendahan Hati	68
I Gusti Ayu Gde Sosiowati	68
Dinamika Bahasa Tanda Luar Ruang Di Lingkungan Sekolah Nasional dan Sekolah internasional Di Wilayah Banjar Badak Sari Denpasar	74
Made Sani Damayanthi Muliawan	74
Pemberdayaan Bahasa Ibu pada Generasi Milenial di Revolusi Industri 4.0: Sebuah Kajian Terhadap Bahasa Minangkabau	82
Oktavianus	82
Kajian Psikolinguistik: Pemerolehan Fonem Bahasa Indonesia pada Anak Usia 4 Tahun (Studi Kasus pada Rafa).....	93
Moh. Yusril Hermansya; Dewi Ayuningtyas.....	93
Cyber Satua: Transformasi Satua Bali Menyambut Revolusi Industri 4.0	99
I Gede Gita Purnama Arsa Putra; Dewa Ayu Carma Citrawati	99
Makna Idiomatis Ungkapan “Mentul-Menceng Mentul-Menceng, Glendang-Glendong Glendang-Glendong” dalam Tradisi Mabuug-Buugan Masyarakat Kedonganan-Badung.....	107
Ida Bagus Rai Putra; Ida Ayu Putu Aridawati	107
Fenomena Bahasa Ibu dalam Dinamika Kemultibahasa dan Keanekabudayaan derta Kemajuan Teknologi Digital: Kasus pada Bahasa Lokal di Pulau Nusa Penida	115

I Ketut Darma Laksana	115
Ketidaksesuaian Ujaran dalam Percakapan Berbahasa Bali terhadap prinsip kooperatif.....	122
I Dewa Ayu Devi Maharani Santika I Gusti Ayu Vina Widiadnya Putri.....	122
Kesantunan Berbahasa Bali Perawat Lansia di Buleleng sebagai Wujud Pemertahanan Bahasa Lokal	131
Putu Dewi Merlyna Yuda Pramesti.....	131
Translating Javanese Lexicons into English in Warung VOA Program of Jawa Timur Television (JTV)	140
Putu Desi Anggerina Hikmaharyanti.....	140
Pencarian Tirta Amerta Dalam Teks Jawa Kuna dan Konteksnya	147
I Made Suastika	147
Analisis Diksi dalam Novel Terjemahan Prancis-Indonesia “Lara Kusapa”	152
Putu Weddha Savitri.....	152
Penanganan Perkara di Desa pada Masa Kerajaan di Bali.....	159
Ida Ayu Wirasmini Sidemen.....	159
Pilihan Berbahasa Generasi Milenial	166
Anak Agung Sagung Shanti Sari Dewi	166
E-learning untuk Mengotimalkan Pembelajaran Bahasa Daerah	173
I Dewa Gede Budi Utama.....	173
Prosedur Penerjemahan Kata Bermuatan Budaya Bali pada Media Cetak Berbahasa Jepang	181
Ladycia Sundayra	181
Pemertahanan Bahasa Bali Melalui Tradisi <i>Nguup</i> di Desa Mundeh Kecamatan Kediri Kabupaten Tabanan	188
Gek Diah Desi Sentana	188
Pemerolehan Bahasa Ibu: Bahasa Bali Anak Usia Dini Pra-Sekolah Melalui Teks <i>Gending Rare</i> Lagu Anak-Anak. Kajian Linguistik Sistemik Fungsional*	195
Putu Utama ¹ ; Maria Arina Luardini ²	195
Permainan Bahasa Humor dalam Akun <i>Selebgram</i> Puja Astawa: Hai Banana.....	201
Komang Dian Puspita Candra ¹ & I Wayan Sidha Karya ²	201
Nomina Bahasa Bali	210
Ni Wayan Suastini	210
Penggunaan Bahasa Bali sebagai Bentuk Pemertahanan Bahasa Ibu pada Media Sosial Instagram	216
Sang Ayu Isnu Maharani	216
Pelanggaran Maksim yang Terjadi dalam Percakapan Puja Astawa.....	224
Ida Bagus Gde Nova Winarta ¹ , Kadek Apriliani ²	224
Explanatory Model dan Rasionalitas Medis Keluarga Orang dengan Gangguan Jiwa (ODGJ) di Desa Blahbatuh Kabupaten Gianyar, Bali.....	234
Bambang Dharwiyanto Putro	234
Struktur Logis Verba Bersufiks {-ang dan -in} Bahasa Bali: Suatu Pendekatan Teori	

Tatabahasa Peran dan Acuan	242
I Nyoman Sedeng	242
Intensitas Konflik Dalam Teks Geguritan <i>Lunga Ka Jembrana</i> Dan Geguritan <i>Mawali Ka Amlapura</i> Karya Anak Agung Istri Agung	249
Tjok. Istri Agung Mulyawati R.	249
Praktik Pendisiplinan Bahasa Ibu Melalui Bahasa Belanda: Pendekakatan Poststrukturalis dalam Sejarah Kebahasaan di Bali	258
I Nyoman Wijaya	258
Tingkat Pemahaman Kosakata Siswa Kelas VI SD dalam Keterampilan Menulis Berbahasa Bali.....	267
Sang Ayu Putu Eny Parwati.....	267
Linguistic Landscapes: Representation of Balinese Letters in Government Building (A Case Study in Gianyar City, Bali).....	275
I Wayan Mulyawan *; Sang Ayu Isnur Maharani; I Gde Budiasa	275
Pemertahanan Bahasa Kaili di Era Milenial melalui Bingkai Sastra	281
Yunidar ¹ ; Ulinsa ²	281
Penggunaan Ungkapan Traditional Bahasa Tae’ dalam Proses Pembelajaran sebagai Wujud Pemertahanan Bahasa Ibu di Sekolah Dasar Kabupaten Luwu Utara	294
Sitti Harisah ¹ ; Ulinsa ²	294
Konteks Tuturan pada Video Humor ‘Turis Apa Turus?’	304
Ni Made Ayu Widiastuti.....	304
Sinergi Antara Industri Kreatif dan Ekonomi Kreatif pada Perajin Bambu dalam Mengatasi Pengangguran di Bali	313
Ketut Darmana	313
Ragam Bahasa Tulis dalam Komunikasi antara Dosen dan Mahasiswa di Prodi Sastra Inggris Universitas Udayana.....	321
Ni Ketut Sri Rahayuni	321
Contrasting the semantics of Indonesian <i>-kan</i> & <i>-i</i> verb pairs: A usage-based, constructional approach.....	328
Gede Primahadi Wijaya Rajeg ¹ ; I Made Rajeg ² ; I Wayan Arka ³	328
Konflik Sosial dan Resolusinya pada Komunitas Bugis di Pulau Serangan Kota Denpasar	345
A.A. Ngurah Anom Kumbara ¹ ; Mayske Rinny Liando ² ; Ida Bagus Gde Putra ³	345
LAMPIRAN NOTULENSI SEMINAR.....	353
Pemakalah: Jufrizal	354
Pemakalah: Ida Ayu Made Puspani.....	355
Pemakalah : Ketut Santi Indriani.....	356
Pemakalah: I Gusti Ngurah Parthama	358
Pemakalah: I Wayan Pastika	360
Pemakalah: Ni Luh Sutjiati Beratha	362
Pemakalah: I Made Sukma Manggala	364

Contrasting the semantics of Indonesian *-kan* & *-i* verb pairs: A usage-based, constructional approach

Gede Primahadi Wijaya Rajeg¹; I Made Rajeg²; I Wayan Arka³
Universitas Udayana^{1,2,3} & Australian National University³
primahadi_wijaya@unud.ac.id¹

Abstract

This paper provides fresh empirical evidence that meanings emerge across different levels of constructions (i.e., morphological and syntactic), and that alternative constructions exhibit complex interaction with grammaticalisation, which possibly alters the meaning and semantic contrast of morphologically related words. To that end, we demonstrate a quantitative method in usage-based, Construction Grammar, namely *Collostructional Analysis*, and apply it in the context of a hypothesis concerning semantic (dis)similarity between verb pairs suffixed with *-kan* and *-i* in Indonesian. We analyse one of the example pairs given in Sneddon et al. (2010), namely *mengenai/mengenakan* (both translated as ‘subject to’ by Sneddon et al.), that they claim to illustrate a blurry semantic difference between *-kan/-i* verbs in common usage. Contrary to Sneddon et al.’s claim, the results from collocational analysis revealed that the verbs convey substantially different meanings. We argue for the importance of quantitative, usage-based approach in addressing hypothesis regarding semantic (dis)similarity of words.

keywords: Indonesian *-kan/-i* verb pairs; Construction Grammar; Usage-based model of language; Quantitative Corpus Linguistics; Collostructional Analysis; R programming language

1 Introduction

Several studies with different theoretical persuasions have attempted to analyse the intricate behaviour of Indonesian verbal suffixes *-kan* and *-i* (Arka, 1993; Arka et al., 2009; Cole & Son, 2004; Kroeger, 2007). Their main themes include the (i) role of the suffixes as valency-changing mechanism, (ii) the resulting argument-structure of the derived verbs, and (iii) the competing account over polysemy/homonymy nature of the suffixes, given their function for causativisation, applicativisation, and transitivisation.

Another area of inquiry is the semantic (dis)similarity between *-kan* and *-i* verb pairs, that is, verb partners with the same root co-occurring with the two suffixes (e.g., *melangkahi* ‘to step over’ and *melangkahkan* ‘to step (the foot) forward’ based on noun root *langkah* ‘step’). When the semantics of the *-kan/-i* verb pairs are contrasted, the focus is mostly on the difference in the semantic roles of their direct objects (cf., e.g., Arka et al., 2009). The Locative/Goal-related roles are mapped onto the direct object of *-i* verbs, while the *-kan*

partners are associated with Non-Locative roles (e.g., Patient, Theme, Beneficiary) (Arka et al., 2009, p. 5; Kroeger, 2007; Sneddon et al., 2010, p. 100; see Rajeg et al., 2019 for a computational and distributional semantic approach with reference to denominal verbs).

One major drawback of the previous works is that argumentation for such semantic difference is typically based on selected usage sentences of a given verb, subjected to grammaticality judgement (but cf. Arka et al., 2009, who also include natural usage data from the web). Such an approach has failed to address, in a more objective and quantifiable manner, the *common usage patterns* of specific *-kan/-i* pairs in large collection of text (i.e., linguistic corpora). The usage patterns we meant here reflect the verbs' co-occurrence distribution with other words. We argue that capturing this co-occurrence distribution of *-kan/-i* verb pairs is by no means (i) less indicative of semantic difference between the contrasted verbs (cf. Stefanowitsch, 2010, p. 368) and (ii) inferior to the mainstream argument-structure and semantic role analyses.

This paper proposes a new perspective in examining the hypothesis regarding semantic difference between *-kan* and *-i* verb pairs (see below), harnessing the growing interest in (i) the usage-based model of language (§1.1) (Diessel, 2017; Goldberg, 2006, pp. 12–14; Langacker, 1988), and (ii) the quantitative turn in linguistics (Janda, 2017). We contextualise this study within the hypotheses that there are *-kan/-i* pairs that exhibit (i) completely different meanings, and/or (ii) indistinguishable meaning (i.e., the pairs are ambiguous in their semantics). We test one of these hypotheses as proposed by Sneddon et al. (2010, p. 101), stating that:

“with a number of words the distinction between **-kan** and **-i** is blurred in *common usage*. In some cases, both **-i** and **-kan** occur with the same meaning. With some there is a recipient or locative object, while with others the object is patient.” (italics ours).

One example given to illustrate this idea is *mengenai/mengenakan* verb-pair, both translated as ‘subject to’ by Sneddon et al. (2010, p. 101). The root of these verbs is *kena* ‘(get) hit (by), knock against; to come into contact with, touch, strike’ (see Stevens & Schmidgall-Tellings, 2004, p. 478). While the hypothesis above is about blurry semantics of *-kan/-i* pairs in “common usage”, Sneddon et al. (2010, p. 101) did not provide any sentential example for how *mengenai* and *mengenakan* come to mean ‘subject to’. Furthermore, the above quote exemplifies a kind of semantic role analysis regarding semantic (dis)similarity between *-kan/-i* pairs.

Our main aim is to look at this long-standing issue in a new light, going beyond investigating the semantic roles and the verbs’ argument-structure pattern. We propose a fully data-driven approach through a quantitative corpus linguistic method, called *Collostructional Analysis* (CollAna) (Stefanowitsch, 2013) (§2.1). CollAna is developed within the spirit of the usage-based, Construction Grammar framework (§1.1) (Goldberg, 2006, 2013). As a preliminary example, we test whether *mengenai* and *mengenakan* do have the same meaning (i.e., ‘subject to’) as Sneddon et al. (2010, p. 101) suggest. As we shall see in §3, collocational data and further corpus-based evidence contradict the semantic-similarity hypothesis between *mengenai/mengenakan*. While this paper only provides analysis for one pair of verbs, the conclusion of which is thus limited, the method is applicable to many other verb pairs.

The remainder of this section provides a brief overview of the usage-based model of language and Construction Grammar (§1.1). We then present the corpus-data source (§2) and describe the *Collostructional Analysis* in more details (§2.1). The results are discussed in §3, followed by the conclusion in §4.

1.1 Usage-based model of language and Construction Grammar

The usage-based model of language (hereafter UBL) is inspired by pioneering research in functional and cognitive linguistics that highlight the centrality of conceptual and pragmatic factors in shaping linguistic structure and meaning (Diessel, 2017). One of the main tenets in UBL is that humans’ linguistic knowledge, including its organisation and structure, is “grounded in the language user’s experience with concrete words and utterances” (Diessel, 2016, p. 209). Another aspect central to UBL is the role of frequency of (co-)occurrence in the structure and meaning of linguistic units (Diessel, 2016, 2017; Diessel & Hilpert, 2016). This paper demonstrates how frequency of co-occurrence between words derived from usage data plays a role in distinguishing meanings of words.

Construction Grammar (CxG) is a theory of linguistic knowledge concerning what speakers know when they need to produce and process a language (cf. Hilpert & Diessel, 2016). CxG and UBL come into close alliance as they share the view that linguistic knowledge (i) is shaped by usage and frequency via domain-general cognitive processes (e.g., automatisisation, schematisation, chunking), and (ii) consists of large interconnected inventory of symbolic units (i.e., form-meaning pairings), called *constructions* (Fillmore, 1988; Goldberg, 2006, 2013). Constructions in CxG range from morpheme, concrete lexical items,

such as words, up to complex phrases, idioms, and abstract grammatical patterns (Fillmore, 1988, p. 42; Goldberg, 2006, 2013).

In generative linguistics, collocation is relegated to a marginal position in linguistic theorising and not part of the theory of grammar (Diessel & Hilpert, 2016). However, in the usage-based framework, such as Construction Grammar, collocation and partially/fully fixed multiword expressions (i.e., prefabricated linguistic units) take a central position in language users’ linguistic knowledge, alongside generalisation of abstract grammatical patterns (Diessel, 2017; Fillmore et al., 1988; Goldberg, 2006, Ch.3; Hilpert & Diessel, 2016, p. 62).

In the usage-based approach to meaning, word co-occurrence, such as collocation, is one operationalisation of *context of use* in characterising the meaning of a given word (Stefanowitsch, 2010). This approach actualises Firth’s (1957, p. 11) famous quote that “you shall know a word by the company it keeps!”, and Wittgenstein’s proposal that “the meaning of a word is its use in the language” (cited from Stefanowitsch, 2010, p. 368). We argue that collocation should not be overlooked in characterising the semantic (dis)similarity of *-kan/-i* verb pairs, given the centrality of collocation as part of language users’ linguistic knowledge as conceived in CxG, and its significance as usage-based evidence for semantics.

2 Data and methods

The analysis in this paper is based on data sourced from one corpus text-file in the *Indonesian Leipzig Corpora* collection (Goldhahn et al., 2012; Quasthoff & Goldhahn, 2013), namely `ind_mixed_2012_1M-sentences.txt`. This file is mostly derived from Indonesian online news website (Quasthoff & Goldhahn, 2013, p. 26) and amounts to 15,052,159 million word-tokens.

2.1 Collostructional Analysis

Collostructional Analysis (CollAna) is a cover term for a family of quantitative corpus linguistic methods that is designed to capture the interface between (abstract) grammatical constructions and lexical items (see Stefanowitsch & Gries, 2003 for the foundational paper; for recent overview, see Stefanowitsch, 2013, 2014). CollAna expands the collocation technique (i.e., word-word co-occurrence) in corpus linguistics into determining the association between word and grammatical constructions.

The earliest iteration of CollAna is called *Simple Collexeme Analysis* (SCA) (Stefanowitsch, 2013). SCA investigates association between a word (i.e., lexical

construction in CxG) and a slot in a grammatical construction (e.g., argument-structure construction). Stefanowitsch and Gries (2003) illustrates SCA to identify verbs strongly attracted to the predicate slot in the English ditransitive construction. Since this construction is not specified lexically, but rather an abstract grammatical pattern for basic sentence type, Stefanowitsch and Gries (2003) propose that the constructional semantics associated with the ditransitive can be read off from the semantics of verbs occurring significantly more frequently in the ditransitive than in other grammatical constructions. It is found that *give* is the most strongly attracted verb in the ditransitive that is semantically compatible with the prototypical ‘transfer’ sense of the ditransitive (see Stefanowitsch & Gries, 2003, pp. 227–230 for details).

Our study extends CollAna to investigate the different usage (i.e. collocation) and probe subtle semantic differences between two morphologically related words²⁵ (Stefanowitsch & Gries, 2003, pp. 217–218). In this paper, we study a pair of verb forms, and, as a preliminary attempt, use their direct object collocates²⁶ for characterising their meanings. However, since the corpus we used is not tagged with any syntactic, morphological, and part-of-speech information, we limit the collocates to be one word immediately following the verb in the corpus (i.e., the verbs’ R1 collocates); hence, the collocational schema of [*mengenai/mengenakan* + w], in which w refers to any ‘word’. We expect that the R1 collocate will capture the direct object slot that immediately follow the verbs (see G. P. W. Rajeg & Rajeg, 2019a for similar approach; cf. Stefanowitsch & Gries, 2003, pp. 220–222, for a study on the word *cause*). This choice is indeed an approximation, and we need to verify that with the results that we obtain.

Table 1 shows the schematic crosstabulation design to calculate the mutual association strength (i.e., degree of collocation strength) between a word and construction (in our case, between a given verb and its R1 collocates) (cf. Stefanowitsch & Gries, 2003, pp. 218–219; Stefanowitsch, 2013).

Table 1: crosstabulation design for calculating association strength between a word and a construction

	word/collocate_X	other_words	Sum
construction/word_X	a	c	a+c

²⁵ Note that in the R programming script (Gries, 2014) to run the variety of CAs, SCA can be optionally used to run collocate analysis.

²⁶ Future work is needed to investigate the nature of the subject NP of the contrasted verbs.

elsewhere	b	d	b+d
Sum	a+b	c+d	a+b+c+d

Cell a in Table 1 represents frequency of co-occurrence between a given construction/word (first row) and its collocate (first column), and is the cell of interest in determining the direction (positive/negative) of association between the construction and its collocate. Cell b is the frequency of the collocate in other contexts and cell c contains the frequency of the construction with other words (cf. Stefanowitsch, 2013). Let us now flesh this out with the frequencies required in SCA to measure the association between *mengenakan* (the target, verbal lexical construction) and *baju* ‘clothes/apparel’ (the R1 collocate of *mengenakan*) (see Table 2); note that the numbers in italics are retrieved directly from the corpus data, the remaining ones are the results of subtractions (Stefanowitsch & Gries, 2003, p. 218).

Table 2: Crosstabulation data for the input of *Fisher-Yates Exact* (FYE) test (numbers in brackets are expected frequencies)

	<i>baju</i>	other_words	Sum
<i>mengenakan</i>	83 (0.15)	1,017 (1,099.85)	<i>1,100</i>
elsewhere	1,552 (1,634.85)	11,884,989 (11,884,906)	11,886,541
Sum	<i>1,635</i>	11,886,006	<i>11,887,641</i>

The direction of association between a given construction and a word is captured by comparing the so-called **observed frequency** ([co-]occurrence frequency in the corpus) with the **expected frequency** ([co-]occurrence frequency expected based on chance alone) (see Gries, 2009, pp. 168–171; G. P. W. Rajeg & Rajeg, 2019b, pp. 21–22, for details on observed vs. expected frequencies). For data in Table 2, we compare the observed and expected frequency in cell a (i.e., the intersection between *mengenakan* and *baju* ‘clothes/apparel’). It turns out that the observed frequency (i.e., 83) is larger than the frequency we would expect by chance when two words co-occur in the corpus (i.e., expected frequency of 0.15). This positive deviation from the expected frequency indicates positive association between *mengenakan* and *baju* (i.e., they are mutually attracted to each other).

The next step is measuring how significantly strong that mutual association is using distributional statistical significance testing. We feed the observed frequencies in Table 2 into the one-tailed *Fisher-Yates Exact* (FYE) test. FYE computes the probability of the observed distribution in Table 2 and all more extreme distributions, maintaining the same marginal sum-frequencies (Stefanowitsch & Gries, 2003, p. 218). The following is a code-snippet for running FYE for Table 2 in R (R Core Team, 2019). We can use the `fisher.test()` function (the

alternative argument is set to "greater" since the observed frequency of cell a is larger than the expected frequency).

```
# create crosstabulation matrix
crosstab <- matrix(data = c(83, 1552, 1017, 11884989), nrow = 2, byrow = FALSE)

# give column and row names
colnames(crosstab) <- c("baju", "other_words")
rownames(crosstab) <- c("mengenakan", "elsewhere")

# Run Fisher-Exact test
pfye <- fisher.test(crosstab, alternative = "greater")

# print out the probability (i.e., p-value)
pfye$p.value

## [1] 9.412147e-196
```

The output p -value above is extremely small ($p < 0.001$)²⁷. It indicates that there is an extremely low probability that the co-occurrence observed frequency of 83 (or even more often) for *mengenakan* and *baju* would have come about by chance alone. In other words, there is a highly significant mutual attraction/association between *mengenakan* and *baju*, such that they can be considered a significantly strong collocation (*mengenakan baju* ‘to wear clothes/apparel’).

For expository reason, the p -value is converted into a base-ten logarithm value (see the R code below), so that positive association (i.e., observed frequency is larger than the expected one) produces positive value and vice versa (see Stefanowitsch & Gries, 2005, p. 7). This value is then called the *collostruction/association strength* (hereafter AssocStr). AssocStr values greater than 1.30103 are significant at the level of $p < 0.05$, while those greater than 2 and 3 are significant at the level of $p < 0.01$ and $p < 0.001$, respectively.

```
# convert into collostruction strength for expository reason
-log(pfye$p.value, 10)

## [1] 195.0263
```

The log-transformed p -value above indicates highly strong attraction between *mengenakan* and *baju* ‘clothes/apparel’. All the above steps need to be repeated to all R1 collocates of the studied verbs; we removed the top-20 most frequent words²⁸ in the corpus from the calculation of the association strength²⁹. Then, the collocates for each verb can be rank

²⁷ In CollAna, threshold for significance is set to be $p < 0.05$.

²⁸ These are the stopwords removed from the analysis: *yang* (n = 505,373); *dan* (n = 363,580); *di* (n = 232,909); *dengan* (n = 182,683); *itu* (n = 172,284); *ini* (n = 149,300); *tidak* (n = 144,373); *untuk* (n = 142,460); *dari* (n = 137,844); *dalam* (n = 129,293); *pada* (n = 91,912); *akan* (n = 89,604); *mereka* (n = 78,934); *kita* (n = 74,574); *ada* (n = 74,210); *saya* (n = 70,730); *karena* (n = 70,469); *adalah* (n = 67,503); *ke* (n = 66,305); *orang* (n = 64,234). They are predominantly function words, such as prepositions, pronouns, conjunctions, determiners.

²⁹ Note that the total size of the corpus (i.e., 11,887,641; see the bottom right cell in Table 2) is reduced from the original size-count (i.e., 15,052,159). This is due to two procedures: (i) the removal of the 20 stop words, and

ordered according to their AssocStr values. To automatise this computation, including the collocates retrieval, we designed an R package called *collogetr* (G. P. W. Rajeg, 2019), and wrote the paper fully in R using R Markdown to interleave the narratives with the R codes (the dataset and R Markdown with codes are available open access as Rajeg et al., 2020).

3 Results and discussion

As mentioned in §1, one issue in Sneddon et al.’s (2010) characterisation for the semantic (dis)similarity between *-kan/-i* verb-pairs, such as *mengenai* and *mengenakan* (both glossed as ‘subject to’), is the absence of actual usage sentences illustrating the meaning of these verbs. Decontextualising the meaning of verbs poses difficulty for foreign learners of Indonesian regarding how the verbs are actually used in sentences. Moreover, Sneddon et al.’s claim that *mengenai* and *mengenakan* convey the same meaning contradicts our native-speaker intuition with how these verbs are actually used. In this section, we present quantitative and usage-based characterisation for *mengenai/mengenakan* pair based on their collocational patterns; we show that these verbs convey different semantics. To begin with, consider Table 3 presenting the top-20 most strongly attracted R1 collocates for *mengenai*.

Table 3: The 20 most strongly attracted R1 collocates for *mengenai*

	collocates	gloss	n	$p_{\text{fisher-exact-value}}$	AssocStr
1	<i>hal</i>	matter	233	3.588e-169	168.445
2	<i>masalah</i>	problem; issue	94	1.407e-77	76.852
3	<i>dampak</i>	impact	46	7.617e-60	59.118
4	<i>keberadaan</i>	existence	35	7.597e-39	38.119
5	<i>hubungan</i>	relationship	49	2.884e-35	34.540
6	<i>apa</i>	what	88	5.698e-35	34.244
7	<i>bagaimana</i>	how	58	1.020e-34	33.992
8	<i>pentingnya</i>	the importance (of)	28	5.077e-34	33.294
9	<i>hal-hal</i>	matters	33	2.498e-29	28.602
10	<i>sasaran</i>	target	22	3.792e-27	26.421
11	<i>rencana</i>	plan	32	1.430e-26	25.845
12	<i>topik</i>	topic	18	1.237e-21	20.908
13	<i>sifat</i>	innate character(istic)	23	1.413e-20	19.850
14	<i>kondisi</i>	condition	31	7.797e-18	17.108

(ii) tokenisation of the original corpus in creating frequency list for collocational analysis; the frequency list only retains alphabetic word-tokens that are at least two-letters long, and, in that case, we removed one-letter tokens, numbers, and incomplete strings (i.e., those initiated or ended with hyphens).

	collocates	gloss	n	$p_{\text{fisher-exact-value}}$	AssocStr
15	<i>karunia-karunia</i>	gifts (of God)	9	2.031e-17	16.692
16	<i>soal</i>	matter; issue	25	3.070e-17	16.513
17	<i>harga</i>	price	27	7.427e-17	16.129
18	<i>siapa</i>	who	30	1.850e-16	15.733
19	<i>keselamatan</i>	safety	16	4.983e-16	15.303
20	<i>adanya</i>	the existence (of)	31	8.566e-16	15.067

The collocates for *mengenai* generally refers to abstract entities, some of which form a coherent semantic class. For instance, *hal* ‘matter’ (see (1)), *hal-hal* ‘matters’, *masalah* ‘issue’ (2), *soal* ‘matter’, and, roughly, *topik* ‘topic’, *dampak* ‘impact’ (3) as well as *rencana* ‘plan’, may represent MATTER-related nouns.

- (1) *Bukti yang paling nyata mengenai hal ini adalah*
evidence REL most real concerning matter DEM is
bagaimana untuk memperebutkan bahan baku dan pasar (...)
how to fight.over raw.material and market
‘The most concrete evidence *regarding* this *matter* is how to fight over the raw materials and markets (...)’ (ind_mixed_2012_1M-sentences.txt:516763)

- (2) *Dan kami berbincang-bincang cukup lama mengenai masalah*
and 1PL.INCL chat enough long (time) concerning problem
itu sampai akhirnya ia kembali menanyakan,
DEM until at.the.end 3SG return ask
“*Ray siapa sih?*”
NAME who tag?
‘And we chatted long enough *regarding* that *problem* until eventually (s)he asks again, “Who is Ray, actually?”’ (ind_mixed_2012_1M-sentences.txt:659875)

- (3) *Dari sumber polutan yang dihasilkan oleh penduduk*
from source pollutant REL PASS.produce by population
diperlukan analisis mengenai dampak negatif terhadap
PASS.need analysis concerning impact negative towards
makhluk hidup di lingkungan tersebut.
living.being LOC environment DEM
‘From the source of pollutants produced by the population, (what is) needed is the analysis *regarding* the negative *impact* (of the pollutants) on the living being in that environment.’
(ind_mixed_2012_1M-sentences.txt:53785)

Two other nouns, namely, *keberadaan* and *adanya*, refer to EXISTENCE. The remaining nouns are rather mixed (e.g., *keselamatan* ‘safety’, *harga* ‘price’, *kondisi* ‘condition’, *pentingnya* ‘the importance (of)’). Another important class of attracted collocates is the WH-words *apa* ‘what’, *bagaimana* ‘how’, and *siapa* ‘who’, introducing the embedded WH-clauses:

- (4) *Mengenai apa yang disampaikan itu menjadi hal berikutnya.*
concerning what REL PASS.deliver=3SG DEM become thing next
'Regarding what (s)he delivered becomes the next matter/thing.' (ind_mixed_2012_1M-sentences.txt:881630)
- (5) *Bahkan, dalam setiap model terdapat ketidakjelasan mengenai bagaimana distribusi otoritas, fungsi, dan hubungan antara institusi-institusi tersebut*
even inside every model there.is unclarity
concerning how distribution authority function and
relationship between institution.PL DEM
'Even, inside every model, there is unclarity regarding how the distribution of authority, function, and relationship between those institutions is.' (ind_mixed_2012_1M-sentences.txt:367454)
- (6) *Ia tidak ingin teman-temannya tahu mengenai siapa 'kakaknya' itu*
3SG NEG want friend.PL know concerning who older.sibling DEM
'(S)he does not want h(is/er) friends know regarding who h(is/er) older sibling is (...)' (ind_mixed_2012_1M-sentences.txt:212649)

The use of *mengenai* with the attracted collocates as illustrated in (1) to (6) demonstrates the grammaticalisation³⁰ of *mengenai*; it exhibits a syntactic function as a connective or clause linker (cf. Hopper & Traugott, 2003, pp. 4, 184), marking an oblique grammatical relation, more like preposition rather than a verbal predicate (see further below). In this function, *mengenai* roughly means 'regarding/concerning/in relation to', rather than 'subject to' (cf. (8) below). Note that this grammaticalised meaning for *mengenai* is listed as its third meaning in *Kamus Besar Bahasa Indonesia* (KBBI) (the Great Dictionary of Indonesian), in addition to its lexical meaning (i.e., 'to come into contact with X') (see (7)).

- (7) *Tak ayal lagi air kotor itu mengenai baju Dimas.*
NEG slow again water dirty DEM come.in.contact shirt NAME
'Soon enough that dirty water hits Dimas' shirt.' (ind_mixed_2012_1M-sentences.txt:774789)

It is argued that the grammatical meaning of *mengenai* (indicating relationship/connection) develops out of its physical meaning (indicating physical contact/touching). That is, when two things are in contact/in touch, they enter some kind of relationship/connection. The results in Table 3 also teach us that the primary meaning of a word in a dictionary may not be the most frequent one in usage as found in a corpus. Here we saw that the grammaticalised meaning of *mengenai* as a connective is much more frequent than its lexical meaning of touching.

³⁰ According to Hopper and Traugott (2003, p. 1), grammaticalisation is a linguistic phenomenon regarding "how lexical items and constructions come in certain linguistic contexts to serve grammatical functions or how grammatical items develop new grammatical functions".

Further evidence of the grammaticalised usage of *mengenai* above, compared to its usage for lexical meaning (as in (7)), is its restricted morphosyntax. That is, *mengenai* meaning ‘concerning/in relation to’ only occurs in active *meN-* form; no paradigmatic opposition/contrast with passive *di-* (i.e. *dikenai*) exists with similar grammaticalised sense. Intuition suggests that the lexical meaning of passive *dikenai* is different from that of the active *mengenai* (exemplified in (7)). Manual scrutiny of 139 usage sentences for *dikenai* in `ind_mixed_2012_1M-sentences.txt` corpus reveals that 89.21% of its usages evoke a situation where the syntactic subject of *dikenai* is *subjected to* certain regulations (e.g., sanction, tax, fee, obligation, punishment, retribution, etc.) (see (8)):

- (8) *Sedangkan motor kedua akan dikenai pajak sebesar 2 persen.*
meanwhile motorbike second FUT PASS.hit tax as.large two percent
‘Meanwhile, the second motorbike will be *subject to/charged with* 2% tax.’
(`ind_mixed_2012_1M-sentences.txt:296558`)

Small percentages of the usage sentences (i.e., around 5.04%) still convey the physical touching/contact meaning (as in (7)):

- (9) *beberapa orang yang dikenai anak panah itu terkapar mati*
several person REL PASS.hit child arrow DEM PASS.sprawled dead
‘Several people who got *hit* by those arrows were sprawled dead’ (`ind_mixed_2012_1M-sentences.txt:81198`)

Brief comparison between the active *mengenai* and passive *dikenai* suggests that the passive form may have different, and conventionalised semantic trait compared to its active version, indicating that a given form (e.g., passive) cannot always be regarded as being derived from the other (i.e., the active). Now let us turn to Table 4 for the attracted collocates of *mengenakan*.

Table 4: The 20 most strongly attracted R1 collocates for *mengenakan*

	collocates	gloss	n	$p_{\text{fisher-exact-value}}$	AssocStr
1	<i>pakaian</i>	clothes	157	0.000e+00	Inf
2	<i>baju</i>	shirt	83	9.412e-196	195.026
3	<i>celana</i>	pants	83	9.948e-195	194.002
4	<i>gaun</i>	dress	37	2.377e-108	107.624
5	<i>jubah</i>	cloak	37	4.368e-103	102.360
6	<i>seragam</i>	uniform	29	3.785e-74	73.422
7	<i>rok</i>	skirt	25	5.774e-65	64.238
8	<i>jilbab</i>	hijab	23	5.276e-63	62.278
9	<i>sepatu</i>	shoes	27	9.158e-59	58.038
10	<i>kaos</i>	T-shirt	21	5.342e-52	51.272

	collocates	gloss	n	$p_{\text{fisher-exact-value}}$	AssocStr
11	<i>busana</i>	clothing; attire	21	6.413e-52	51.193
12	<i>jaket</i>	jacket	16	9.825e-43	42.008
13	<i>jas</i>	jacket; coat	15	2.033e-40	39.692
14	<i>topi</i>	hat; cap	13	1.383e-33	32.859
15	<i>kerudung</i>	veil; hood	11	8.588e-32	31.066
16	<i>bra</i>	bra	12	4.933e-31	30.307
17	<i>kostum</i>	costume	11	4.658e-30	29.332
18	<i>helm</i>	helmet	12	1.644e-27	26.784
19	<i>kemeja</i>	(western-style) shirt	10	1.448e-25	24.839
20	<i>sarung</i>	sarong	11	7.382e-25	24.132

As shown in Table 4, it is clear that *mengenakan* exhibits substantially different and rather restricted collocational patterns compared to *mengenai*. In this top-20 list, *mengenakan* is exclusively associated with CLOTHING and ACCESSORIES nouns, and conveys a lexical meaning of ‘to put on/wear X (where X is usually clothing or accessories)’:

- (10) *Siswi yang masih mengenakan seragam putih abu-abu itu melemparkan senyuman ke arah kendaraan.*
 student.FEM REL still wear uniform white grey DEM
 throw smile to direction vehicle
 ‘That (female) student who is still *wearing* (her) white-and-grey *uniforms* casts a smile towards the vehicle.’ (ind_mixed_2012_1M-sentences.txt:722930)

Another crucial difference between *mengenakan* and *mengenai* is that the former (i.e., the *-kan* form) never occurs in a more abstract, grammaticalised context (i.e., ‘concerning’) as in the latter. This is in line with the grammaticalisation patterns attested in other languages where logical relation, such as CONCERN, has been typically developed out of a word/formative marker carrying locative meaning (see Heine & Kuteva, 2002, pp. 201-202), and previous studies in Indonesian linguistics (e.g., Arka et al., 2009) have shown that the suffix *-i* (but not *-kan*) carries the locative meaning. In this respect, in relation to the tested hypothesis (§1), if Sneddon et al.’s (2010, p. 101) intention for the semantic similarity between *mengenai* and *mengenakan* refers to the grammaticalised meaning of *mengenai* discussed above, then, *mengenakan* is not the substitute for *mengenai* in that grammaticalised sense, for the reason that *mengenai* and *mengenakan* are semantically distinct in their usage patterns. This claim is supported by the results of corpus-based, collocational analyses for the two words above.

4 Conclusion

This study set out to demonstrate how a hypothesis concerning words meanings can be tested in quantitative terms. The hypothesis that we tested is one of the classic issues in Indonesian linguistics, namely, blurry semantic distinction between some of the verb pairs suffixed with *-kan* and *-i* (Sneddon et al., 2010, p. 101; cf. Arka et al., 2009). Previous studies of this issue have generally been restricted to the analysis of semantic roles of the direct object of the contrasted verbs and largely ignored the *actual* usage patterns of the contrasted verbs in large collection of text (i.e., linguistic corpora). Leveraging the availability of large language corpora, the advent of quantitative corpus-linguistic method, and usage-based linguistics, we propose the application of *Collostructional Analysis* (CollAna) (Stefanowitsch, 2013) (§2.1) in addressing that long-standing issue in a new light. As a case study, we analysed a verb-pair whose characterisation is problematic from our perspective as Indonesian native-speakers, namely *mengenai/mengenakan*; both are translated as ‘subject to’ by Sneddon et al. (2010, p. 101), but without any sentential examples provided.

Contrary to Sneddon et al’s claim for the verbs’ similar meaning, our findings clearly indicate that *mengenai* and *mengenakan* exhibit distinct collocational preferences, showing substantial semantic differences between them. *Mengenakan*’s top-20 attracted collocates predominantly refer to CLOTHING words (Table 4), evoking the meaning of ‘to wear X (clothes)’. In contrast, *mengenai* attracts abstract nouns (Table 3), and is used predominantly as a grammaticalised connective, expressing ‘regarding/concerning/in relation to’ sense (cf. example (1) to (6)). The findings reported here shed new light on the value of keeping an eye on actual usage data and applying quantitative method in (i) characterising semantic (dis)similarity between words, as in the case of the *-kan/-i* verb pairs in Indonesian, and (ii) finding highly conventionalised verb-collocate pairings.

5 References

- Arka, I. W. (1993). *Morpholexical aspects of the -kan causative in Indonesian* [Master Thesis]. University of Sydney.
- Arka, I. W., Dalrymple, M., Mistica, M., Mofu, S., Andrews, A. D., & Simpson, J. (2009). A linguistic and computational morphosyntactic analysis for the applicative *-i* in Indonesian. In M. Butt & T. H. King (Eds.), *Proceedings of the LFG09 conference*. CSLI Publications.
- Cole, P., & Son, M.-J. (2004). The argument structure of verbs with the suffix *-kan* in Indonesian. *Oceanic Linguistics*, 43(2), 339–364.

- Diessel, H. (2016). Frequency and lexical specificity in grammar: A critical review. In H. Behrens & S. Pfänder (Eds.), *Experience counts: Frequency effects in language*. De Gruyter. <https://doi.org/10.1515/9783110346916-009>
- Diessel, H. (2017). Usage-based linguistics. In M. Aronoff (Ed.), *Oxford research encyclopedia of linguistics*. Oxford University Press. <https://doi.org/10.1093/acrefore/9780199384655.013.363>
- Diessel, H., & Hilpert, M. (2016). Frequency effects in grammar. In M. Aronoff (Ed.), *Oxford research encyclopedia of linguistics*. Oxford University Press. <https://doi.org/10.1093/acrefore/9780199384655.013.120>
- Fillmore, C. J. (1988). The mechanisms of "construction grammar". *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society*, 14, 35–55. <https://doi.org/10.3765/bls.v14i0.1794>
- Fillmore, C. J., Kay, P., & O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language*, 64(3), 501. <https://doi.org/10.2307/414531>
- Firth, J. R. (1957). *A synopsis of linguistic theory 1930-55. 1952-59*, 1–32. <http://cs.brown.edu/courses/csci2952d/readings/lecture1-firth.pdf>
- Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press.
- Goldberg, A. E. (2013). Constructionist approaches. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford handbook of construction grammar* (pp. 15–31). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195396683.013.0002>
- Goldhahn, D., Eckart, T., & Quasthoff, U. (2012). Building large monolingual dictionaries at the Leipzig corpora collection: From 100 to 200 languages. *Proceedings of the 8th Language Resources and Evaluation Conference (LREC) 2012*, 759–765. http://www.lrec-conf.org/proceedings/lrec2012/pdf/327_Paper.pdf
- Gries, S. T. (2009). *Statistics for linguistics with R: A practical introduction*. Mouton de Gruyter.
- Gries, S. T. (2014). *Coll.analysis 3.5. A script for r to compute perform collostructional analysis*. <http://www.linguistics.ucsb.edu/faculty/stgries/teaching/groningen/readme.txt>
- Heine, B., & Kuteva, T. (2002). *World lexicon of grammaticalization*. Cambridge University Press.
- Hilpert, M., & Diessel, H. (2016). Entrenchment in construction grammar. In H.-J. Schmid (Ed.), *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge* (pp. 57–74). De Gruyter. <https://doi.org/10.1515/9783110341423-004>
- Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization* (2nd ed.). Cambridge University Press.
- Janda, L. A. (2017). The quantitative turn. In B. Dancygier (Ed.), *The Cambridge handbook of cognitive linguistics* (pp. 498–514). Cambridge University Press. <https://doi.org/10.1017/9781316339732.032>

- Kroeger, P. (2007). Morphosyntactic vs. Morphosemantic functions of Indonesian *-kan*. In A. Zaenen (Ed.), *Architectures, rules, and preferences: Variations on themes by Joan W. Bresnan* (1 edition, pp. 229–251). Center for the Study of Language; Information.
- Langacker, R. W. (1988). A usage-based model. In B. Rudzka-Ostyn (Ed.), *Topics in cognitive linguistics* (pp. 127–161). John Benjamins Publishing Company.
- Quasthoff, U., & Goldhahn, D. (2013). *Indonesian corpora* (Nos. 7). Abteilung Automatische Sprachverarbeitung, Institut für Informatik, Universität Leipzig.
<http://asvdoku.informatik.uni-leipzig.de/corpora/data/uploads/corpus-building-vol7-ind.pdf>
- Rajeg, G. P. W. (2019). *Collogetr: Collocates retriever and collocation association measure* (Version 1.1.3) [Computer software]. <https://doi.org/10.26180/5b7b9c5e32779>
- Rajeg, G. P. W., & Rajeg, I. M. (2019a). Analisis koleksem khas dan potensinya untuk kajian kemiripan makna konstruksional dalam bahasa Indonesia. In I. N. Sudipa (Ed.), *ETIKA BAHASA buku persembahan menapaki usia pensiun: I Ketut Tika* (Vol. 1, pp. 65–83). Swasta Nulus. <https://doi.org/10.26180/5bf4e49ea1582>
- Rajeg, G. P. W., & Rajeg, I. M. (2019b). 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, G. P. W., Denistia, K., & Musgrave, S. (2019). Vector Space Models and the usage patterns of Indonesian denominal verbs: A case study of verbs with meN-, meN-/kan, and meN-/i affixes. (H. Nomoto & D. Moeljadi, Eds.) *NUSA, Linguistic studies using large annotated corpora*, 67, 35–76. <http://repository.tufs.ac.jp/handle/10108/94452>
- Rajeg, G. P. W., Rajeg, I. M., & Arka, I. W. (2020). *R Markdown Notebook and data for “Contrasting the semantics of Indonesian -kan and -i verb pairs: A usage-based, constructional approach”*. <https://doi.org/10.6084/m9.figshare.11823366.v1>
- R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Sneddon, J. N., Adelaar, A., Djenar, D. N., & Ewing, M. C. (2010). *Indonesian reference grammar* (2nd ed.). Allen & Unwin.
- Stefanowitsch, A. (2014). Collostructional analysis: A case study of the English *into-causative*. In T. Herbst, H.-J. Schmid, & S. Faulhaber (Eds.), *Constructions collocations patterns* (pp. 217–238). Walter De Gruyter, GmbH.
- Stefanowitsch, A. (2010). Empirical cognitive semantics: Some thoughts. In Dylan Glynn & Kerstin Fischer (Eds.), *Quantitative methods in cognitive semantics: Corpus-driven approaches* (pp. 355–380). Mouton de Gruyter.
- Stefanowitsch, A. (2013). Collostructional analysis. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford handbook of construction grammar* (pp. 290–306). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195396683.013.0016>
- Stefanowitsch, A., & Gries, S. T. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8(2), 209–243.
- Stefanowitsch, A., & Gries, S. T. (2005). Covarying collexemes. *Corpus Linguistics and Linguistic Theory*, 1(1), 1–43.

Stevens, A. M., & Schmidgall-Tellings, A. E. (2004). *A comprehensive Indonesian-English dictionary*. Ohio University Press.