

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

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

The slides are available open-access at:  
**<http://bit.ly/snbi-collana>**

# Our aims

- Demonstrate a quantitative method in addressing a long-standing theoretical issue in Indonesian linguistics:
  - Semantic (dis)similarity of *–kan/-i* verb-pairs

# Our aims

- Demonstrate a quantitative method in addressing a long-standing theoretical issue in Indonesian linguistics:
  - Semantic (dis)similarity of *–kan/-i* verb-pairs
- Provide fresh, empirical (i.e. usage-based) and quantitative evidence:
  - meanings emerge at different constructional levels (i.e., morphological and syntactical)
    - Different meaning is associated with distinct morphology and collocational patterns
  - alternating constructions exhibit complex interaction with grammaticalisation

# Background

- *-kan* and *-i* suffixes have received serious attention in Indonesian (theoretical) linguistics (e.g., Arka, 1993; Arka et al., 2009; Cole & Son, 2004; Kroeger, 2007)

- 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.
- 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 and Information.

# Background

- *-kan* and *-i* suffixes have received serious attention in Indonesian (theoretical) linguistics (e.g., Arka, 1993; Arka et al., 2009; Cole & Son, 2004; Kroeger, 2007)
- Main themes are on:
  - the role of the suffixes as valency-changing mechanism
  - the resulting argument-structure of the derived verbs
  - the debate over *-kan* and *-i* complex functions (i.e. causative/applicative polysemy/homonymy)

- 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.
- 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 and Information.

# Issues

- Previous studies are mostly **inexplicit** about their source of evidence (but cf. Arka et al. 2009)
  - Argumentation is typically based on grammaticality judgement of (introspectively) constructed linguistic examples

# Issues

- Previous studies are mostly **inexplicit** about their source of evidence (but cf. Arka et al. 2009)
  - Argumentation is typically based on grammaticality judgement of (introspectively) constructed linguistic examples
- Semantic distinction analysis between *–kan/–i* verb-pairs is limited:
  - Mainly focusing on the semantic roles of the direct object (e.g., locative vs non-locative)
  - Based on small number of sentential examples

# Issues

- Previous studies are mostly **inexplicit** about their source of evidence (but cf. Arka et al. 2009)
  - Argumentation is typically based on grammaticality judgement of (introspectively) constructed linguistic examples
- Semantic distinction analysis between *–kan/-i* verb-pairs is limited:
  - Mainly focusing on the semantic roles of the direct object (e.g., locative vs non-locative)
  - Based on small number of sentential examples
- Little is known about how specific verbs with both suffixes are **actually used in large collection of text (linguistic corpus)**

# Looking at old issue in a new light...

- Semantic (dis)similarity between *–kan/–i* verb-pairs
- Test the related hypothesis using a quantitative method in usage-based, Construction Grammar (Goldberg, 2006, 2013):
  - COLLOSTRUCTIONAL ANALYSIS (Stefanowitsch, 2013)

- 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>
- 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>

# Hypothesis

“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.” (Sneddon et al., 2010, p. 101; italics ours)

# Hypothesis

“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.” (Sneddon et al., 2010, p. 101; italics ours)

- One example given by Sneddon et al. (2010) is ***mengenai/mengenakan***
  - both translated as ‘subject to’ by Sneddon et al. (2010)
  - **HOWEVER**, no “*common usage*” sentences are given for these two words!

# Hypothesis

“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.” (Sneddon et al., 2010, p. 101; italics ours)

- One example given by Sneddon et al. (2010) is ***mengenai/mengenakan***
  - both translated as ‘subject to’ by Sneddon et al. (2010)
  - **HOWEVER**, no “*common usage*” sentences are given for these two words!
- We’ll demonstrate a case study on how such hypothesis can be tested in **quantitative** terms, following the framework of, and technique in, usage-based, Construction Grammar

# Construction Grammar

“A theory of what speakers know when they know how to produce and process language” (Hilpert, 2019)

- Hilpert, M. (2019). *What is Construction Grammar?*. Presented as part of the Lecture Series on *Ten Lectures on Diachronic Construction Grammar*

# Construction Grammar (CxG)

(Fillmore et al., 1988; Goldberg, 2006)

- No strict distinction between lexicon and syntax
- Non-derivational: “What you see is what you get” (Goldberg, 2006, p. 10)
- No strict distinction between competence and performance:
  - Our **knowledge of language** (i) is **emergent bottom-up** from language use, and (ii) feeds into further use of that knowledge (Bybee, 2013; Hopper, 1987)

- Bybee, J. L. (2013). Usage-based theory and exemplar representations of constructions. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford handbook of Construction Grammar* (pp. 49–69). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195396683.013.0004>
- 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>
- Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press.
- Hopper, P. J. (1987). Emergent Grammar. *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society*, 139–157. <https://doi.org/10.3765/bls.v13i0.1834>

# Construction Grammar (CxG)

(Fillmore et al., 1988; Goldberg, 2006)

- **All knowledge of language consists of a network of learned pairing of form and meaning/function, that is CONSTRUCTION** (Goldberg, 2006)
  - “constructions all the way down” (Goldberg, 2006, p. 18)
- Construction varies in their symbolic complexity and schematicity (Langacker, 2013; Croft, 2001; Goldberg, 2006):
  - from morpheme, to simple word, idiomatic phrases, and up to (partially) schematic syntactic constructions

- Croft, W. (2001). *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford University Press.
- 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>
- Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press.
- Langacker, R. W. (2013). *Essentials of Cognitive Grammar*. Oxford University Press.

# Construction Grammar (CxG)

TABLE 1.1. Examples of constructions, varying in size and complexity

---

|                                 |                                                                                          |
|---------------------------------|------------------------------------------------------------------------------------------|
| Morpheme                        | e.g. <i>pre-</i> , <i>-ing</i>                                                           |
| Word                            | e.g. <i>avocado</i> , <i>anaconda</i> , <i>and</i>                                       |
| Complex word                    | e.g. <i>daredevil</i> , <i>shoo-in</i>                                                   |
| Complex word (partially filled) | e.g. [N-s] (for regular plurals)                                                         |
| Idiom (filled)                  | e.g. <i>going great guns</i> , <i>give the Devil his due</i>                             |
| Idiom (partially filled)        | e.g. <i>jog</i> <someone's> <i>memory</i> , <i>send</i> <someone> <i>to the cleaners</i> |
| Covariational Conditional       | The Xer the Yer (e.g. <i>the more you think about it, the less you understand</i> )      |
| Ditransitive (double object)    | Subj V Obj1 Obj2 (e.g. <i>he gave her a fish taco</i> ; <i>he baked her a muffin</i> )   |
| Passive                         | Subj aux VPpp (PP <sub>by</sub> ) (e.g. <i>the armadillo was hit by a car</i> )          |

---

(Goldberg, 2006, p. 5)

- Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press.

# Construction Grammar (CxG)

- **Collocational preference** is part of speakers' **linguistic knowledge** (Hilpert & Diessel, 2016, p. 62)
  - It is assumed that speakers stored information about collocation (i.e. associative links between frequently co-occurring linguistic units)
- In Usage-based Linguistics, collocation is a method to define “context of use” in characterising meanings of (lexical and grammatical) constructions (Stefanowitsch, 2010)
  - **OUR STUDY** adopts the collocational approach in CxG (i.e. **COLLOSTRUCTIONAL ANALYSIS**)

- 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>
- 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.

# Collostructional Analysis

Collocate analysis in a Constructional framework

# Collostructional Analysis

(Stefanowitsch & Gries, 2003, 2009)

- An analytical method in Quantitative Corpus Linguistics couched within the usage-based Construction Grammar
- CollAna determines the most typical collocates of a construction

- Hilpert, M. (2014). Collostructional analysis: Measuring associations between constructions and lexical elements. In D. Glynn & J. A. Robinson (Eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy* (pp. 391–404). John Benjamins Publishing Company.
- 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. Th. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8(2), 209–243.
- Stefanowitsch, A., & Gries, S. Th. (2009). Corpora and grammar. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: An international handbook* (Vol. 2, pp. 933–951). Mouton de Gruyter.

# Collostructional Analysis

(Stefanowitsch & Gries, 2003, 2009)

- An analytical method in Quantitative Corpus Linguistics couched within the usage-based Construction Grammar
- CollAna determines the most typical collocates of a construction
- **Collocational preferences** of constructions may **reveal the meanings of the constructions**

- Hilpert, M. (2014). Collostructional analysis: Measuring associations between constructions and lexical elements. In D. Glynn & J. A. Robinson (Eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy* (pp. 391–404). John Benjamins Publishing Company.
- 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. Th. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8(2), 209–243.
- Stefanowitsch, A., & Gries, S. Th. (2009). Corpora and grammar. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: An international handbook* (Vol. 2, pp. 933–951). Mouton de Gruyter.

# Collostructional Analysis

(Stefanowitsch & Gries, 2003, 2009)

- Simple Collexeme/Collocates Analysis (SCA)
  - typical collocates of a construction
  - E.g., typical verbs in ditransitive construction (see Stefanowitsch, 2013)

- Hilpert, M. (2014). Collostructional analysis: Measuring associations between constructions and lexical elements. In D. Glynn & J. A. Robinson (Eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy* (pp. 391–404). John Benjamins Publishing Company.
- 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. Th. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8(2), 209–243.
- Stefanowitsch, A., & Gries, S. Th. (2009). Corpora and grammar. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: An international handbook* (Vol. 2, pp. 933–951). Mouton de Gruyter.

# Collostructional Analysis

(Stefanowitsch & Gries, 2003, 2009)

- Simple Collexeme/Collocates Analysis (SCA)
  - typical collocates of a construction
  - E.g., typical verbs in ditransitive construction (see Stefanowitsch, 2013)
- **Application of SCA in this study:**
  - the construction: “lexical constructions” with *-kan* and *-i* suffixes
  - the collocates: one word to the right of the verbs (R1 collocates)
    - R1 collocates seek to approximate the fillers of direct object (DO) of the verbs
  - **Semantic types of the DO may reveal the semantics of the verbs**

- Hilpert, M. (2014). Collostructional analysis: Measuring associations between constructions and lexical elements. In D. Glynn & J. A. Robinson (Eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy* (pp. 391–404). John Benjamins Publishing Company.
- 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. Th. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8(2), 209–243.
- Stefanowitsch, A., & Gries, S. Th. (2009). Corpora and grammar. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: An international handbook* (Vol. 2, pp. 933–951). Mouton de Gruyter.

# Brief notes on data and method

- **One corpus-text file** from the *Indonesian Leipzig Corpora Collection* (Biemann et al., 2007; Goldhahn et al., 2012; Quasthoff & Goldhahn, 2013)
  - `ind_mixed_2012_1M-sentences.txt` (1 Mill. Sentences; 15,052,159 Mill word-tokens)
  - The corpus is raw (no annotation/tagging for Part-of-Speech and Syntactic Dependency)
- Collocates retrieval and analysis with R package **collogetr** (Rajeg, 2019)

- Biemann, C., Heyer, G., Quasthoff, U., & Richter, M. (2007). The Leipzig Corpora Collection: Monolingual corpora of standard size. In M. Davies, P. Rayson, S. Hunston, & P. Danielsson (Eds.), *Proceedings of the Corpus Linguistics Conference*. [http://ucrel.lancs.ac.uk/publications/CL2007/paper/190\\_Paper.pdf](http://ucrel.lancs.ac.uk/publications/CL2007/paper/190_Paper.pdf)
- 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](http://www.lrec-conf.org/proceedings/lrec2012/pdf/327_Paper.pdf)
- Quasthoff, U., & Goldhahn, D. (2013). *Indonesian corpora* (No. 7; Technical Report Series on Corpus Building). Abteilung Automatische Sprachverarbeitung, Institut für Informatik, Universität Leipzig. <http://asvdoku.informatik.uni-leipzig.de/corpora/data/uploads/corpus-building-vol7-ind.pdf>
- R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rajeg, G. P. W. (2019). *collogetr: Collocates retriever and Collocation association measure* (Version 1.1.3) [R]. <https://doi.org/10.26180/5b7b9c5e32779>

# Simple Collexeme/Collocates Analysis

The diagram illustrates a Simple Collexeme/Collocates Analysis table. The table has four columns: 'Target (lexical) construction', 'Target (R1) collocates', 'Other collocates', and 'row totals'. The rows are 'mengenakan', 'elsewhere', and 'row totals'. Annotations include red arrows pointing from labels to specific cells: 'Target (lexical) construction' points to the first column, 'Other contexts' points to the 'elsewhere' row, 'Target (R1) collocates' points to the 'baju' cell, and 'Other collocates' points to the '¬ baju' cell. The cells 'mengenakan', 'baju', and '¬ baju' are circled in red.

| Target (lexical) construction | Target (R1) collocates | Other collocates | row totals |
|-------------------------------|------------------------|------------------|------------|
| mengenakan                    | baju                   | ¬ baju           |            |
| elsewhere                     |                        |                  |            |
| row totals                    |                        |                  |            |

# Simple Collexeme/Collocates Analysis

Freq. of collocation in the  
collocation window  
(*mengenakan* + *baju*)

All token freq. of  
*mengenakan* in the  
corpus

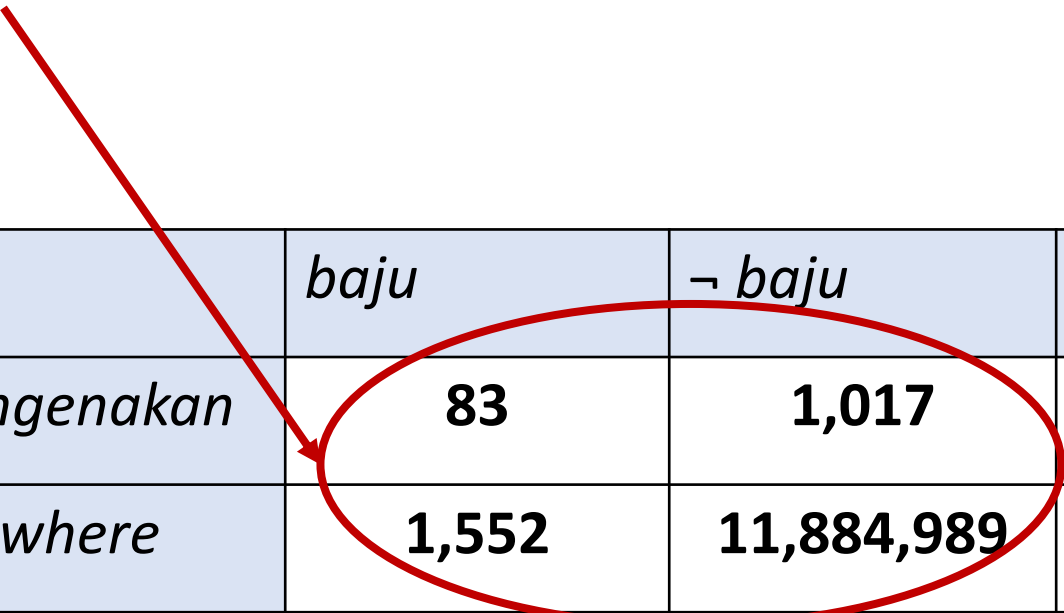
|                   | <i>baju</i> | ¬ <i>baju</i> | row totals |
|-------------------|-------------|---------------|------------|
| <i>mengenakan</i> | 83          |               | 1,100      |
| <i>elsewhere</i>  |             |               |            |
| row totals        | 1,635       |               | 11,887,641 |

All token freq. of *baju*  
in the corpus

Number of  
word-tokens  
in the corpus  
(i.e. corpus  
size)

# Simple Collexeme/Collocates Analysis

**Compute Association Strength  
(with Fisher-Yates Exact test)**



|                   | <i>baju</i>  | $\neg$ <i>baju</i> | row totals |
|-------------------|--------------|--------------------|------------|
| <i>mengenakan</i> | <b>83</b>    | <b>1,017</b>       | 1,100      |
| <i>elsewhere</i>  | <b>1,552</b> | <b>11,884,989</b>  | 11,886,541 |
| row totals        | 1,635        | 11,886,006         | 11,887,641 |

# Simple Collexeme/Collocates Analysis

**Compute Association Strength  
(with Fisher-Yates Exact test)**

|                   | <i>baju</i>  | $\neg$ <i>baju</i> | row totals |
|-------------------|--------------|--------------------|------------|
| <i>mengenakan</i> | <b>83</b>    | <b>1,017</b>       | 1,100      |
| <i>elsewhere</i>  | <b>1,552</b> | <b>11,884,989</b>  | 11,886,541 |
| row totals        | 1,635        | 11,886,006         | 11,887,641 |

# Simple Collexeme/Collocates Analysis



```
# create 2-by-2 crosstabulation matrix
crosstab <- matrix(crosstab <- matrix(data = c(83, 1552, 1017, 11884989), nrow = 2, byrow = FALSE),
  dimnames = list(Lexical_CxN = c("mengenakan", "elsewhere"),
    Collocates = c("baju", "-baju")))
```

```
# print out the crosstabulation
```

```
crosstab
```

```
#           Collocates
# Lexical_CxN baju    -baju
# mengenakan    83     1017
# elsewhere    1552    11884989
```

|            | baju  | - baju     |
|------------|-------|------------|
| mengenakan | 83    | 1,017      |
| elsewhere  | 1,552 | 11,884,989 |

```
# Run Fisher-Exact test and get the p-value as measure of association strength
```

```
pfye <- fisher.test(crosstab, alternative = "greater")$p.value
```

```
# print out the p-value, which is very small
```

```
# indicating low likelihood that we observe a random distribution
```

```
# i.e. there's strong mutual association between "mengenakan" and "baju"
```

```
pfye
```

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

```
# convert into collostruction strength for expository reason
```

```
-log(pfye, 10)
```

```
# [1] 195.0263 <- Collostruction Strength (very strong association at  $p < 0.001$ )
```

- R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rajeg, G. P. W. (2019). *collogetr: Collocates retriever and Collocation association measure* (Version 1.1.3) [R]. <https://doi.org/10.26180/5b7b9c5e32779>

# Results & Discussion

## The top-20 most strongly attracted R1 collocates for *mengenakan*

|    | <b>collocates</b> | <b>gloss</b>          | <b>n</b> | <b>p<sub>fisher-exact</sub>-value</b> | <b>AssocStr</b> |
|----|-------------------|-----------------------|----------|---------------------------------------|-----------------|
| 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          |
| 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          |

## The top-20 most strongly attracted R1 collocates for *mengenakan*

|    | <b>collocates</b> | <b>gloss</b>          | <b>n</b> | <b>p<sub>fisher-exact</sub>-value</b> | <b>AssocStr</b> |
|----|-------------------|-----------------------|----------|---------------------------------------|-----------------|
| 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                 |          |                                       |                 |
| 4  | <i>gaun</i>       | dress                 |          |                                       |                 |
| 5  | <i>jubah</i>      | cloak                 |          |                                       |                 |
| 6  | <i>seragam</i>    | uniform               |          |                                       |                 |
| 7  | <i>rok</i>        | skirt                 |          |                                       |                 |
| 8  | <i>jilbab</i>     | hijab                 |          |                                       |                 |
| 9  | <i>sepatu</i>     | shoes                 |          |                                       |                 |
| 10 | <i>kaos</i>       | T-shirt               |          |                                       |                 |
| 11 | <i>busana</i>     | clothing; attire      |          |                                       |                 |
| 12 | <i>jaket</i>      | jacket                |          |                                       |                 |
| 13 | <i>jas</i>        | jacket; coat          |          |                                       |                 |
| 14 | <i>topi</i>       | hat; cap              |          |                                       |                 |
| 15 | <i>kerudung</i>   | veil; hood            |          |                                       |                 |
| 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          |

*Mengenakan* expresses  
'to wear (clothes)'

Attracted nouns for  
**CLOTHING and BODY-RELATED ACCESSORIES**

- (11) *Siswi*                    *yang*    *masih*    ***mengenakan***    ***seragam***    *putih*    *abu-abu*    *itu*  
student.FEM    REL    still    wear                    uniform    white    grey    DEM  
*melemparkan*            *senyuman*    *ke*    *arah*                    *kendaraan.*  
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)

## The top-20 most strongly attracted R1 collocates for *mengenai*

|    | <b>collocates</b>      | <b>gloss</b>            | <b>n</b> | <b>p<sub>fisher-exact</sub>-value</b> | <b>AssocStr</b> |
|----|------------------------|-------------------------|----------|---------------------------------------|-----------------|
| 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          |
| 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 top-20 most strongly attracted R1 collocates for *mengenai*

|    | collocates             | gloss                   | n   | p <sub>fisher-exact</sub> -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                  |     |                                  |          |
| 4  | <i>keberadaan</i>      | existence               |     |                                  |          |
| 5  | <i>hubungan</i>        | relationship            |     |                                  |          |
| 6  | <i>apa</i>             | what                    |     |                                  |          |
| 7  | <i>bagaimana</i>       | how                     |     |                                  |          |
| 8  | <i>pentingnya</i>      | the importance (of)     |     |                                  |          |
| 9  | <i>hal-hal</i>         | matters                 |     |                                  |          |
| 10 | <i>sasaran</i>         | target                  |     |                                  |          |
| 11 | <i>rencana</i>         | plan                    |     |                                  |          |
| 12 | <i>topik</i>           | topic                   |     |                                  |          |
| 13 | <i>sifat</i>           | innate character(istic) |     |                                  |          |
| 14 | <i>kondisi</i>         | condition               |     |                                  |          |
| 15 | <i>karunia-karunia</i> | gifts (of God)          |     |                                  |          |
| 16 | <i>soal</i>            | matter; issue           |     |                                  |          |
| 17 | <i>harga</i>           | price                   |     |                                  |          |
| 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   |

*Mengenai* expresses  
'concerning/in relation to'

Reflecting the  
predominantly  
grammaticalised usage as  
a connective

## The top-20 most strongly attracted R1 collocates for *mengenai*

|    | collocates      | gloss                   | n   | p <sub>fisher-exact</sub> -value | AssocStr |
|----|-----------------|-------------------------|-----|----------------------------------|----------|
| 1  | hal             | matter                  | 233 | 3.588e-169                       | 168.445  |
| 2  | masalah         | problem; issue          | 94  | 1.407e-77                        | 76.852   |
| 3  | dampak          | impact                  |     |                                  |          |
| 4  | keberadaan      | existence               |     |                                  |          |
| 5  | hubungan        | relationship            |     |                                  |          |
| 6  | apa             | what                    |     |                                  |          |
| 7  | bagaimana       | how                     |     |                                  |          |
| 8  | pentingnya      | the importance (of)     |     |                                  |          |
| 9  | hal-hal         | matters                 |     |                                  |          |
| 10 | sasaran         | target                  |     |                                  |          |
| 11 | rencana         | plan                    |     |                                  |          |
| 12 | topik           | topic                   |     |                                  |          |
| 13 | sifat           | innate character(istic) |     |                                  |          |
| 14 | kondisi         | condition               |     |                                  |          |
| 15 | karunia-karunia | gifts (of God)          |     |                                  |          |
| 16 | soal            | matter; issue           | 25  | 3.070e-17                        | 16.513   |
| 17 | harga           | price                   | 27  | 7.427e-17                        | 16.129   |
| 18 | siapa           | who                     | 30  | 1.850e-16                        | 15.733   |
| 19 | keselamatan     | safety                  | 16  | 4.983e-16                        | 15.303   |
| 20 | adanya          | the existence (of)      | 31  | 8.566e-16                        | 15.067   |

*Mengenai* expresses  
'concerning/in relation to'

Attracting abstract nouns,  
emphasising *mengenai*'s  
abstract meaning

# Grammaticalised usage of *mengenai*

|                      |                 |                        |                      |                   |                 |                 |
|----------------------|-----------------|------------------------|----------------------|-------------------|-----------------|-----------------|
| <i>Dari</i>          | <i>sumber</i>   | <i>polutan</i>         | <i>yang</i>          | <i>dihasilkan</i> | <i>oleh</i>     | <i>penduduk</i> |
| from                 | source          | pollutant              | REL                  | PASS.produce      | by              | population      |
| <i>diperlukan</i>    | <i>analisis</i> | <b><i>mengenai</i></b> | <b><i>dampak</i></b> | <i>negatif</i>    | <i>terhadap</i> |                 |
| PASS.need            | analysis        | concerning             | impact               | negative          | towards         |                 |
| <i>makhluk hidup</i> | <i>di</i>       | <i>lingkungan</i>      | <i>tersebut.</i>     |                   |                 |                 |
| 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)

# Grammaticalised usage of *mengenai*

|                      |                 |                        |                      |                   |                 |                 |
|----------------------|-----------------|------------------------|----------------------|-------------------|-----------------|-----------------|
| <i>Dari</i>          | <i>sumber</i>   | <i>polutan</i>         | <i>yang</i>          | <i>dihasilkan</i> | <i>oleh</i>     | <i>penduduk</i> |
| from                 | source          | pollutant              | REL                  | PASS.produce      | by              | population      |
| <i>diperlukan</i>    | <i>analisis</i> | <b><i>mengenai</i></b> | <b><i>dampak</i></b> | <i>negatif</i>    | <i>terhadap</i> |                 |
| PASS.need            | analysis        | concerning             | impact               | negative          | towards         |                 |
| <i>makhluk hidup</i> | <i>di</i>       | <i>lingkungan</i>      | <i>tersebut.</i>     |                   |                 |                 |
| 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)

Grammaticalisation of *mengenai* (with Locative *–i*) offers further evidence to the tendency of LOCATIVE meaning to develop into grammatical marker indicating logical relation, such as CONCERN (see Heine & Kuteva (2002, p. 201-202)).

# Grammaticalised usage of *mengenai*

|                      |                 |                        |                      |                   |                 |                 |
|----------------------|-----------------|------------------------|----------------------|-------------------|-----------------|-----------------|
| <i>Dari</i>          | <i>sumber</i>   | <i>polutan</i>         | <i>yang</i>          | <i>dihasilkan</i> | <i>oleh</i>     | <i>penduduk</i> |
| from                 | source          | pollutant              | REL                  | PASS.produce      | by              | population      |
| <i>diperlukan</i>    | <i>analisis</i> | <b><i>mengenai</i></b> | <b><i>dampak</i></b> | <i>negatif</i>    | <i>terhadap</i> |                 |
| PASS.need            | analysis        | concerning             | impact               | negative          | towards         |                 |
| <i>makhluk hidup</i> | <i>di</i>       | <i>lingkungan</i>      | <i>tersebut.</i>     |                   |                 |                 |
| 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)

## Constrained morphosyntax:

- NO PARADIGMATIC OPPOSITION (e.g., *dikenai*) for this grammaticalised meaning
- NO grammaticalised sense for the *-kan* form (i.e., *mengenakan* ‘to wear’)

## The top-20 most strongly attracted R1 collocates for *mengenai*

|    | collocates             | gloss                   | n   | p <sub>fisher-exact</sub> -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                  |     |                                  |          |
| 4  | <i>keberadaan</i>      | existence               |     |                                  |          |
| 5  | <i>hubungan</i>        | relationship            |     |                                  |          |
| 6  | <i>apa</i>             | what                    |     |                                  |          |
| 7  | <i>bagaimana</i>       | how                     |     |                                  |          |
| 8  | <i>pentingnya</i>      | the importance (of)     |     |                                  |          |
| 9  | <i>hal-hal</i>         | matters                 |     |                                  |          |
| 10 | <i>sasaran</i>         | target                  |     |                                  |          |
| 11 | <i>rencana</i>         | plan                    |     |                                  |          |
| 12 | <i>topik</i>           | topic                   |     |                                  |          |
| 13 | <i>sifat</i>           | innate character(istic) |     |                                  |          |
| 14 | <i>kondisi</i>         | condition               |     |                                  |          |
| 15 | <i>karunia-karunia</i> | gifts (of God)          |     |                                  |          |
| 16 | <i>soal</i>            | matter; issue           |     |                                  |          |
| 17 | <i>harga</i>           | price                   |     |                                  |          |
| 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   |

*Mengenai* expresses  
'concerning/in relation to'

Co-occurring with WH-  
word reflecting co-  
occurrence with  
embedded clause

# Grammaticalised usage of *mengenai*

- (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)

# Lexical usage of *mengenai* means ‘to hit sth.’

*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)

# Discussion

- *Mengenakan* is not the substitute for *mengenai* in the sense of 'concerning/in relation to':
  - They have different semantic traits (e.g., distinct collocational preferences)
  - They also differ in their syntactic function:
    - *Mengenai* is now grammaticalised into an oblique marker (like preposition)
      - No evidence for paradigmatic opposition (i.e. *dikenai*) expressing this grammaticalised sense
    - *Mengenakan* is used predominantly as a lexical verb meaning 'to wear'

# Discussion

- SCA provides one usage-based, quantitative evidence **against** Sneddon et al's (2010) hypothesis that *mengenai/mengenakan* exemplify words reflecting semantic similarity of *-i/-kan* verb pairs
  - They **do not** convey the **same meaning!**
- Similar method can be applied to other verb pairs (e.g., *menanami* vs. *menanamkan*)

# Desiderata

- Widen the collocate (e.g., the subject NP)
- Investigate other morphological variants based on *kena* ‘(get) hit (by)’ (e.g., *dikenakan*, *dikenai*).
- More research on similar pairs

# Thank you!

**Questions? Comments? All welcome!**

The slides are available open-access at:  
**<http://bit.ly/snbi-collana>**