

Distinctive Collexeme Analysis of Indonesian Causative Rival Affixes *per-* and *-kan*

Gede Primahadi Wijaya Rajeg ^a & Karlina Denistia ^b

Universitas Udayana ^a & Universitas Sebelas Maret ^b

primahadi_wijaya@unud.ac.id ^a & karlinadenistia@staff.uns.ac.id ^b

Presented at the plenary session of the KOLITA 19 on the 15th of July 2021

How to cite materials from this presentation:

Rajeg, Gede Primahadi Wijaya; Denistia, Karlina (2021): Distinctive Collexeme Analysis of Indonesian Causative Rival Affixes *per-* and *-kan*. figshare.

Conference contribution presented at the KOLITA 19.

<https://doi.org/10.6084/m9.figshare.13602155>

1. Rival affixes
2. *per-* and *-kan*
3. Distinctive Collexeme Analysis
4. Results
5. Discussion & Conclusions

1. Rival affixes 

2. *per-* and *-kan*

3. Distinctive Collexeme Analysis

4. Results

5. Discussion & Conclusions

Rival affixes

Different forms, but similar meaning and fulfilling similar derivational and syntactic functions (Aronoff and Anshen 2017)

Rival affixes

Different forms, but similar meaning and fulfilling similar derivational and syntactic functions (Aronoff and Anshen 2017)

(1)	academic-ize	*academic-ify
	computer-ize	*computer-ate
	art-ify	*art-ate

Adopted from Plag 1999

Rival affixes

Different forms, but similar meaning and fulfilling similar derivational and syntactic functions (Aronoff and Anshen 2017)

(1)	academic-ize	*academic-ify
	computer-ize	*computer-ate
	art-ify	*art-ate

Adopted from Plag 1999

- Quantitative analysis on Indonesian rival affixes
 - Agent with PE- & PEN- (Denistia and Baayen 2019, Denistia et.al. 2021)
 - Applicative -KAN and -I (Rajeg, Denistia & Musgrave 2019)

~~1. Rival affixes~~

2. *per-* and *-kan* 

3. Distinctive Collexeme Analysis

4. Results

5. Discussion & Conclusions

per- and *-kan* in the literature

- *per-* is historically older
 - Reflex of PAN causative prefix **pa-* (Adelaar 1992: 49; Ogloblin 1998: 182)
 - *-kan* is thought to develop from **(a)kAn* 'towards' (Ogloblin 1998: 182)

per- and *-kan* in the literature

- *per-* is historically older
 - Reflex of PAN causative prefix **pa-* (Adelaar 1992: 49; Ogloblin 1998: 182)
 - *-kan* is thought to develop from **(a)kAn* ‘towards’ (Ogloblin 1998: 182)
- *per-* vs. *-kan* is less studied
 - Semantic ambiguity/generalisation of *per-* and *-kan*
 - “many people making no distinction between the function of the two affixes” (Sneddon et al. 2010: 103)
 - Collocational & productivity constraints for *per-* and *-kan* (Rajeg & Rajeg 2019, Rajeg & Denistia 2021)

per- and *-kan*

Sneddon et al. (2010) indicated the **base constraints**, “with some bases, one affix or the other does not occur or is very infrequent”

base word	<i>per-</i>	<i>-kan</i>
<i>besar</i> ‘big’	<i>memperbesar</i> ‘to make something bigger’	<i>membesarkan</i> ‘to make something bigger’
<i>tetap</i> ‘fix’	-	<i>menetapkan</i> ‘to make something fixed’
<i>cantik</i> ‘pretty’	<i>mempercantik</i> ‘to make something/someone prettier’	<i>mencantikkan</i> ‘to make something/someone prettier’
<i>ganteng</i> ‘handsome’	<i>memperganteng</i> ‘to make something/someone more handsome’	-

per- and *-kan*

Sneddon et al. (2010) indicated the **base constraints**, “with some bases, one affix or the other does not occur or is very infrequent”

base word	<i>per-</i>	<i>-kan</i>
<i>besar</i> ‘big’	<i>memperbesar</i> ‘to make something bigger’	<i>membesarkan</i> ‘to make something bigger’
<i>tetap</i> ‘fix’	-	<i>menetapkan</i> ‘to make something fixed’
<i>cantik</i> ‘pretty’	<i>mempercantik</i> ‘to make something/someone prettier’	<i>mencantikkan</i> ‘to make something/someone prettier’
<i>ganteng</i> ‘handsome’	<i>memperganteng</i> ‘to make something/someone more handsome’	-

per- and *-kan*

Sneddon et al. (2010) indicated the **base constraints**, “with some bases, one affix or the other does not occur or is very infrequent”

base word	<i>per-</i>	<i>-kan</i>
<i>besar</i> ‘big’	<i>memperbesar</i> ‘to make something bigger’	<i>membesarkan</i> ‘to make something bigger’
<i>tetap</i> ‘fix’	-	<i>menetapkan</i> ‘to make something fixed’
<i>cantik</i> ‘pretty’	<i>mempercantik</i> ‘to make something/someone prettier’	<i>mencantikkan</i> ‘to make something/someone prettier’
<i>ganteng</i> ‘handsome’	<i>memperganteng</i> ‘to make something/someone more handsome’	-

Research questions

1. What are the adjectival bases that exhibit asymmetry for *per*- and *-kan*?
2. What semantic preferences can be explored from the list of the distinctive adjectival bases for *per*- and *-kan*?

~~1. Rival affixes~~

~~2. *per-* and *-kan*~~

3. Distinctive Collexeme Analysis 

4. Results

5. Discussion & Conclusions

Distinctive Collexeme Analysis (DCA) - *What is it?*

- Differences between two **functionally/semantically similar** constructions (CxNs) in terms of statistical association of lexical elements that can occur with the CxNs

Distinctive Collexeme Analysis (DCA) - *What is it?*

- Differences between two **functionally/semantically similar** constructions (CxNs) in terms of statistical association of lexical elements that can occur with the CxNs



<https://imgflip.com/i/3pnzw9>

Table 5. Collexemes distinguishing between the *will* and the *be going to* future

WILL (N=3,667)		BE GOING TO (N=980)	
Collexeme	Distinctiveness	Collexeme	Distinctiveness
see (90:8)	0.0004	say (28:42)	1.12E-12
find (58:4)	0.0015	do (105:68)	2.02E-08
give (69:7)	0.0047	happen (11:15)	4.77E-05
know (34:2)	0.0108	have (126:61)	0.0001
provide (17:0)	0.0177	go (96:48)	0.0004
depend (15:0)	0.0285	win (1:6)	0.0005
want (22:1)	0.0305	stay (8:10)	0.0014
receive (14:0)	0.0361	use (19:14)	0.0045
consider (13:0)	0.0458	buy (2:5)	0.0059
remain (13:0)	0.0458	talk (16:11)	0.0160
become (19:1)	0.0553	show (22:13)	0.0213
finish (12:0)	0.0581	get (84:34)	0.0275
hold (11:0)	0.0736	suggest (1:3)	0.0315
include (11:0)	0.0736	be (664:203)	0.0357
notice (11:0)	0.0736	put (19:11)	0.0362
follow (10:0)	0.0934	invest (0:2)	0.0444
reach (10:0)	0.0934	measure (0:2)	0.0444
need (16:1)	0.0985	perform (0:2)	0.0444
send (21:2)	0.1080	photocopy (0:2)	0.0444
accept (9:0)	0.1184	rehearse (0:2)	0.0444

Distinctiveness represents the p -value (p) of the statistical significance test with Fisher-Yates Exact (FYE) test.

The smaller the p -value (i.e. the smaller p than 0.05 [$p < 0.05$]), the more distinctive a given collexeme is to the respective construction.

p -value is used to rank-order the collexeme

Table 5. Collexemes distinguishing between the *will* and the *be going to* future

WILL (N=3,667)		BE GOING TO (N=980)	
<i>Collexeme</i>	<i>Distinctiveness</i>	<i>Collexeme</i>	<i>Distinctiveness</i>
see (90:8)	0.0004	say (28:42)	1.12E-12
find (58:4)	0.0015	do (105:68)	2.02E-08
give (69:7)	0.0047	happen (11:15)	4.77E-05
know (34:2)	0.0108	have (126:61)	0.0001
provide (17:0)	0.0177	go (96:48)	0.0004
depend (15:0)	0.0285	win (1:6)	0.0005
want (22:1)	0.0305	stay (8:10)	0.0014
receive (14:0)	0.0361	use (19:14)	0.0045
consider (13:0)	0.0458	buy (2:5)	0.0059
remain (13:0)	0.0458	talk (16:11)	0.0160
become (19:1)	0.0553	show (22:13)	0.0213
finish (12:0)	0.0581	get (84:34)	0.0275
hold (11:0)	0.0736	suggest (1:3)	0.0315
include (11:0)	0.0736	be (664:203)	0.0357
notice (11:0)	0.0736	put (19:11)	0.0362
follow (10:0)	0.0934	invest (0:2)	0.0444
reach (10:0)	0.0934	measure (0:2)	0.0444
need (16:1)	0.0985	perform (0:2)	0.0444
send (21:2)	0.1080	photocopy (0:2)	0.0444
accept (9:0)	0.1184	rehearse (0:2)	0.0444

“Dynamicity of the actions and events encoded” (Gries & Stefanowitsch 2004: 114)

Table 5. Collexemes distinguishing between the *will* and the *be going to* future

WILL (N=3,667)		BE GOING TO (N=980)	
<i>Collexeme</i>	<i>Distinctiveness</i>	<i>Collexeme</i>	<i>Distinctiveness</i>
see (90:8)	0.0004	say (28:42)	1.12E-12
find (58:4)	0.0015	do (105:68)	2.02E-08
give (69:7)	0.0047	happen (11:15)	4.77E-05
know (34:2)	0.0108	have (126:61)	0.0001
provide (17:0)	0.0177	go (96:48)	0.0004
depend (15:0)	0.0285	win (1:6)	0.0005
want (22:1)	0.0305	stay (8:10)	0.0014
receive (14:0)	0.0361	use (19:14)	0.0045
consider (13:0)	0.0458	buy (2:5)	0.0059
remain (13:0)	0.0458	talk (16:11)	0.0160
become (19:1)	0.0553	show (22:13)	0.0213
finish (12:0)	0.0581	get (84:34)	0.0275
hold (11:0)	0.0736	suggest (1:3)	0.0315
include (11:0)	0.0736	be (664:203)	0.0357
notice (11:0)	0.0736	put (19:11)	0.0362
follow (10:0)	0.0934	invest (0:2)	0.0444
reach (10:0)	0.0934	measure (0:2)	0.0444
need (16:1)	0.0985	perform (0:2)	0.0444
send (21:2)	0.1080	photocopy (0:2)	0.0444
accept (9:0)	0.1184	rehearse (0:2)	0.0444

“Dynamicity of the actions and events encoded” (Gries & Stefanowitsch 2004: 114)

Distinctive collexemes for WILL are predominantly relatively “non-agentive or low-dynamicity actions” (e.g., *find, receive, hold, finish, reach*)

perception/cognition events (*see, know, want, consider, notice, need, accept*)

states (*depend, remain, become*).

Five collexemes encode dynamic actions: *give, provide, include, follow, send*

Table 5. Collexemes distinguishing between the *will* and the *be going to* future

WILL (N=3,667)		BE GOING TO (N=980)	
<i>Collexeme</i>	<i>Distinctiveness</i>	<i>Collexeme</i>	<i>Distinctiveness</i>
see (90:8)	0.0004	say (28:42)	1.12E-12
find (58:4)	0.0015	do (105:68)	2.02E-08
give (69:7)	0.0047	happen (11:15)	4.77E-05
know (34:2)	0.0108	have (126:61)	0.0001
provide (17:0)	0.0177	go (96:48)	0.0004
depend (15:0)	0.0285	win (1:6)	0.0005
want (22:1)	0.0305	stay (8:10)	0.0014
receive (14:0)	0.0361	use (19:14)	0.0045
consider (13:0)	0.0458	buy (2:5)	0.0059
remain (13:0)	0.0458	talk (16:11)	0.0160
become (19:1)	0.0553	show (22:13)	0.0213
finish (12:0)	0.0581	get (84:34)	0.0275
hold (11:0)	0.0736	suggest (1:3)	0.0315
include (11:0)	0.0736	be (664:203)	0.0357
notice (11:0)	0.0736	put (19:11)	0.0362
follow (10:0)	0.0934	invest (0:2)	0.0444
reach (10:0)	0.0934	measure (0:2)	0.0444
need (16:1)	0.0985	perform (0:2)	0.0444
send (21:2)	0.1080	photocopy (0:2)	0.0444
accept (9:0)	0.1184	rehearse (0:2)	0.0444

“Dynamicity of the actions and events encoded” (Gries & Stefanowitsch 2004: 114)

Distinctive collexemes for *BE going to*, however, predominantly encode dynamic actions.

Only five collexemes encode states or non-agentive actions (*have, stay, be, happen, get*)

Another difference:

- **Specificity of the actions and events**
 - ***invest, measure, photocopy***

Distinctive Collexeme Analysis (DCA) - *What is it?*

- Differences between two **functionally/semantically similar** constructions (CxNs) in terms of statistical association of lexical elements that can occur with the CxNs
- **Syntactic alternations**
 - Active vs. Passive
 - Ditransitive vs. *to*-Dative
 - Genitive *of* vs. *'s*



<https://imgflip.com/i/3pnzw9>

Distinctive Collexeme Analysis (DCA) - *What is it?*

- Differences between two **functionally/semantically similar** constructions (CxNs) in terms of statistical association of lexical elements that can occur with the CxNs



<https://imgflip.com/i/3pnzw9>

Extension into morphological phenomena in Indonesian

[*per*+ADJ] vs. [ADJ+*kan*]

Why and in what ways is DCA relevant to causative *per*- and -*kan*?

per- and *-kan*

Sneddon et al. (2010) indicated the **base constraints**, “with some bases, one affix or the other does not occur or is very infrequent”

base word	<i>per-</i>	<i>-kan</i>
besar `big`	memperbesar `to make something bigger`	membesarkan `to make something bigger`
tetap `fix`	-	menetapkan `to make something fixed`
cantik `pretty`	mempercantik `to make something/someone prettier`	mencantikkan `to make something/someone prettier`
ganteng `handsome`	memperganteng `to make something/someone more handsome`	-

Other methodological aspects

- Thirteen files of the *Indonesian Leipzig Corpora* (180,769,204 word-tokens)
- Target verbal forms:
 - *meN-* (**mem***per*besar, *membesar***kan**)
 - *di-* (**di***per*besar, *dibesar***kan**)
 - unprefix imperative (*per*besar, *besar***kan**)
- *MorphInd* (Larasati, Kubon & Zeman 2011)
 - Morphological parsing, incl. word-class tagging of the base
- *MALINDO Morph* (Nomoto et al. 2018) & KBBI
 - Manual verification of the *MorphInd* output
 - Manual verification for the hapaxes (one-token occurrence of the verbal form)

**Total 842 adjectival
base types occurring
with *PER-* and *-KAN***

Distinctive Collexeme Analysis (DCA) - *How does it work?*

CONSTRUCTION variable
(with two levels: CxN A & CxN B)

Collexemes	Construction (CxN) A	Construction (CxN) B	Row Total
Collexeme L	Freq. of L in A	Freq. of L in B	Total Freq. of L
Collexemes other than L	Freq. of others in A	Freq. of others in B	Total Freq of Others
Column Total	Freq. of CxN A	Freq of. CxN B	Sum of CxN A & CxN B

Distinctive Collexeme Analysis (DCA) - *How does it work?*

[*per*+ADJ] vs. [ADJ+*kan*]

	<i>per</i> +ADJ	ADJ+ <i>kan</i>	Row Total
Adjective L	Freq. of L in <i>per</i> +ADJ	Freq. of L in ADJ+ <i>kan</i>	Total Freq. of L
Adjectives other than L	Freq. of other adjectives in <i>per</i> +ADJ	Freq. of other adjectives in ADJ+ <i>kan</i>	Total Freq of other adjectives
Column Total	Total Freq. of <i>per</i> +ADJ	Total Freq. of ADJ+ <i>kan</i>	Sum of <i>per</i> +ADJ & ADJ+ <i>kan</i>

Distinctive Collexeme Analysis (DCA) - *How does it work?*

[*per*+ADJ] vs. [ADJ+*kan*]

	<i>per</i> +ADJ	ADJ+ <i>kan</i>	Row Total
<i>kaya</i>	2,628	34	2,662
Other adjectives	76,268	820,336	896,604
Column Total	78,896	820,370	899,266

Distinctive Collexeme Analysis (DCA) - *How does it work?*

[*per*+ADJ] vs. [ADJ+*kan*]

	<i>per</i> +ADJ	ADJ+ <i>kan</i>	Row Total
<i>kaya</i>	2,628	34	2,662
Other adjectives	76,268	820,336	896,604
Column Total	78,896	820,370	899,266

Input for the distributional statistical test (i.e., significance test) with Fisher-Yates Exact Test

Distinctive Collexeme Analysis (DCA) - *How does it work?*


[*per*+ADJ] vs. [ADJ+*kan*]

Compute p-value of Fisher-Yates Exact test -> Collostruction Strength (-log10)

```
> fisher.test(tbl, alternative = "greater")
```



Fisher's Exact Test for Count Data

```
data:  tbl
p-value < 2.2e-16
alternative hypothesis: true odds ratio is greater than 1
95 percent confidence interval:
 612.5714      Inf
sample estimates:
odds ratio
 829.8198
```



***kaya* is a highly significantly distinctive adjectival base for one of the CxNs**

```
> fisher.test(tbl, alternative = "greater")$p.value
[1] 0
> -log10(fisher.test(tbl, alternative = "greater")$p.value)
[1] Inf
```



Distinctive Collexeme Analysis (DCA) - *How does it work?*

[*per*+ADJ] vs. [ADJ+*kan*]

Compare Obs. vs. Exp. Freq.

PER-kaya is more frequent than expected.
kaya-KAN is less frequent than expected.

	<i>per</i> +ADJ	ADJ+ <i>kan</i>	Row Total
<i>kaya</i>	O > E 2,628 (exp: 234) +	O < E 34 (exp: 2,428) -	2,662
Other adjectives	76,268 (exp: 78,662)	820,336 (exp: 817,942)	896,604
Column Total	78,896	820,370	899,266


Kaya is highly significantly distinctive for PER- than -KAN.
 In other words, the form *perkaya* is a highly significant (& potentially cognitively strongly entrenched) morphological construct than *kayakan*.

Distinctive Collexeme Analysis (DCA) - *studies in Indonesian & Balinese*

Explanation (in both English and Indonesian), R codes, and application of DCA on some topics of Indonesian and Balinese linguistics can be seen at **Rajeg, Denistia, & Rajeg (2018); Rajeg & Rajeg (2019; see 2021 for Balinese data)**

- ~~1. Rival affixes~~
- ~~2. *per-* and *-kan*~~
- ~~3. Distinctive Collexeme Analysis~~
4. Results 
5. Discussion & Conclusions

Research questions

1. What are the adjectival bases that exhibit asymmetry for *per*- and *-kan*? 

Top-20 Distinctive Adjectival Bases for -KAN

	ADJ khas untuk -kan	N _{per-}	N _{-kan}	CollStr
1	<i>nyata</i>	3	96.957	Inf
2	<i>jelas</i>	1.298	72.444	Inf
3	<i>tentu</i>	0	45.182	Inf
4	<i>tetap</i>	0	41.289	Inf
5	<i>pasti</i>	0	37.570	Inf
6	<i>siap</i>	13	28.088	Inf
7	<i>umum</i>	11	21.487	Inf
8	<i>selamat</i>	0	16.156	Inf
9	<i>sesuai</i>	0	14.020	Inf
10	<i>aman</i>	0	13.543	Inf

Negative evidence:

“constructions that **did not occur** and **could not have** (these can be referred to as ‘significantly absent’ structures)” (Stefanowitsch 2006: 62, our boldface)

“with some bases, one affix or the other **does not occur** or is very infrequent” (Sneddon et al. 2010: 103)

Top-20 Distinctive Adjectival Bases for -KAN

Top-20 Distinctive Adjectival Bases for

	ADJ khas untuk -kan	N _{per-}	N _{-kan}	CollStr
1	<i>nyata</i>	3	96.957	Inf
2	<i>jelas</i>			
	ADJ khas untuk -kan	N _{per-}	N _{-kan}	CollStr
3	<i>tentu</i>	11	<i>rugi</i>	013.301Inf
4	<i>tetap</i>	12	<i>bersih</i>	112.653Inf
5	<i>pasti</i>	13	<i>senang</i>	011.709Inf
6	<i>siap</i>	14	<i>bebas</i>	110.350Inf
7	<i>umum</i>	15	<i>beda</i>	09.821Inf
8	<i>selamat</i>	16	<i>khawatir</i>	08.809Inf
9	<i>sesuai</i>	17	<i>raya</i>	18.698Inf
10	<i>aman</i>	18	<i>benar</i>	08.604Inf
		19	<i>batal</i>	08.556Inf
		20	<i>yakin</i>	17.763308,1463

Highlights - negative evidence

- 203 (79.3%) out of the total 256 distinctive adjectival bases for *-kan* are significantly absent with *per-* (i.e., occurring with the frequency of zero with *per-*)

Highlights - negative evidence

- 203 (79.3%) out of the total 256 distinctive adjectival bases for *-kan* are significantly absent with *per-* (i.e., occurring with the frequency of zero with *per-*)
- 12 (19.05%) out of the total 63 distinctive adjectival bases for *per-* are significantly absent with *-kan* (i.e., occurring with the frequency of zero with *-kan*)

ADJ khas untuk *per-* N_{per-} N_{-kan} CollStr

<i>kuat</i>	18.957	4.643	Inf
<i>cepat</i>	9.196	1	Inf
<i>luas</i>	7.451	334	Inf
<i>panjang</i>	6.281	322	Inf
<i>mudah</i>	4.870	9.547	Inf
<i>banyak</i>	2.693	9	Inf
<i>kaya</i>	2.628	34	Inf
<i>besar</i>	2.584	3.993	Inf
<i>kecil</i>	1.907	786	Inf
<i>ketat</i>	1.795	118	Inf

Top-20 Distinctive Adjectival Bases for *PER-*

Raw obs. Freq.
are higher for
-KAN but still
proportionally
more frequent for
PER-

The adjectives occur with
BOTH *PER-* and *-KAN* but are
proportionally asymmetrical:
they are significantly more
frequent than expected with
PER-

ADJ khas untuk <i>per-</i>	N _{<i>per-</i>}	N _{<i>-kan</i>}	CollStr
----------------------------	--------------------------	--------------------------	---------

<i>kuat</i>	18.957	4.643	Inf
-------------	--------	-------	-----

ADJ khas untuk <i>per-</i>	N _{<i>per-</i>}	N _{<i>-kan</i>}	CollStr
----------------------------	--------------------------	--------------------------	---------

<i>luas</i>	<i>parah</i>	1.666	2	Inf
-------------	--------------	-------	---	-----

<i>panjang</i>	<i>lambat</i>	1.656	134	Inf
----------------	---------------	-------	-----	-----

<i>mudah</i>	<i>sulit</i>	1.407	2.549	Inf
--------------	--------------	-------	-------	-----

<i>banyak</i>	<i>dalam</i>	1.267	14	Inf
---------------	--------------	-------	----	-----

<i>kaya</i>	<i>buruk</i>	1.232	34	Inf
-------------	--------------	-------	----	-----

<i>besar</i>	<i>sempit</i>	1.065	121	Inf
--------------	---------------	-------	-----	-----

<i>kecil</i>	<i>cantik</i>	979	4	Inf
--------------	---------------	-----	---	-----

<i>ketat</i>	<i>kokoh</i>	840	399	Inf
--------------	--------------	-----	-----	-----


	<i>tajam</i>	783	208	Inf
--	--------------	-----	-----	-----

	<i>pendek</i>	658	109	Inf
--	---------------	-----	-----	-----

Top-20 Distinctive Adjectival Bases for *PER-*

Quantified operationalisation for
the **base constraints** hypothesis for
per+ADJ and ADJ+*kan*

Research questions

- ~~1. What are the adjectival bases that exhibit asymmetry for *per-* and *-kan*?~~
2. What semantic preferences can be explored from the list of the distinctive adjectival bases for *per-* and *-kan*? 

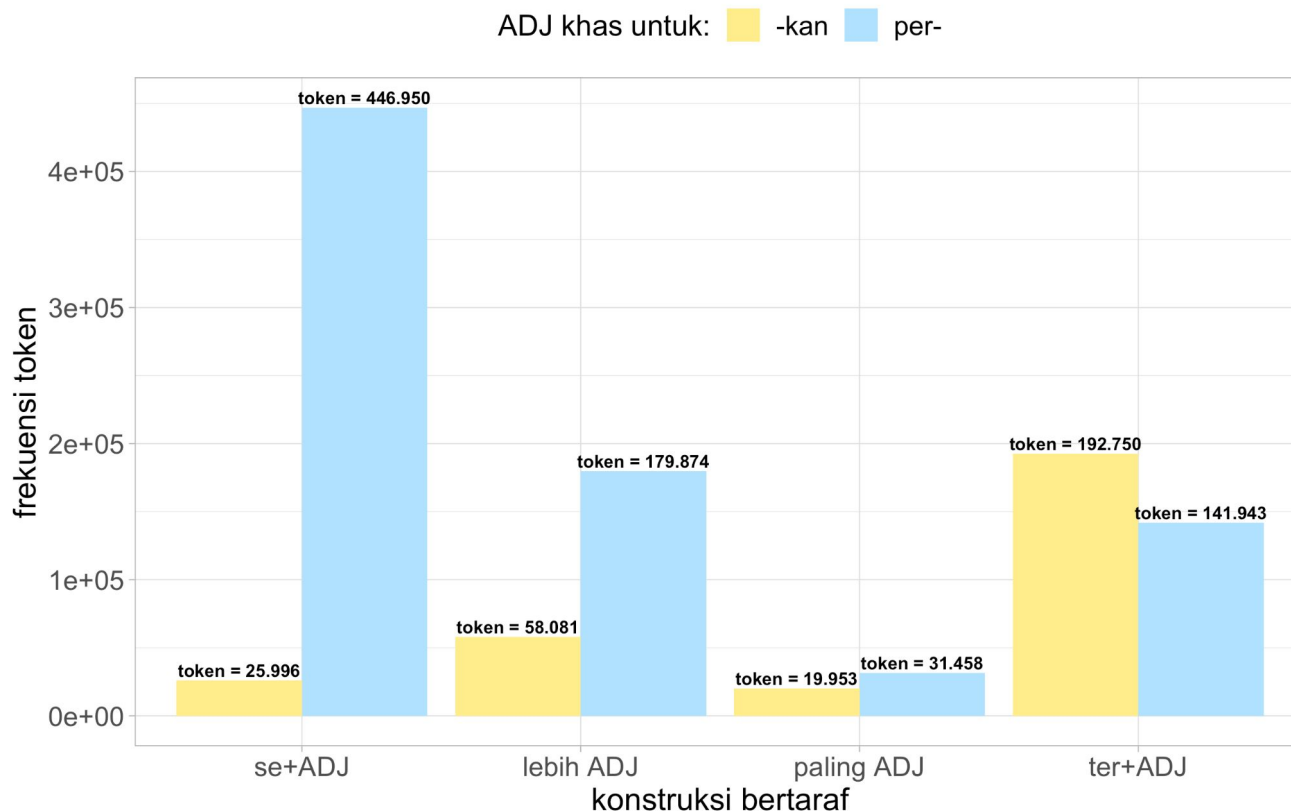
Semantic characteristics of the distinctive adjectives for *per-* and *-kan*

From the top-20 list

Adjective	<i>per-</i>	<i>-kan</i>	Examples
emotion/psych.			<i>yakin, senang, siap, khawatir</i>
spatial / space / size / dimension			<i>besar, dalam, kecil, luas, panjang, pendek, sempit, ketat</i>

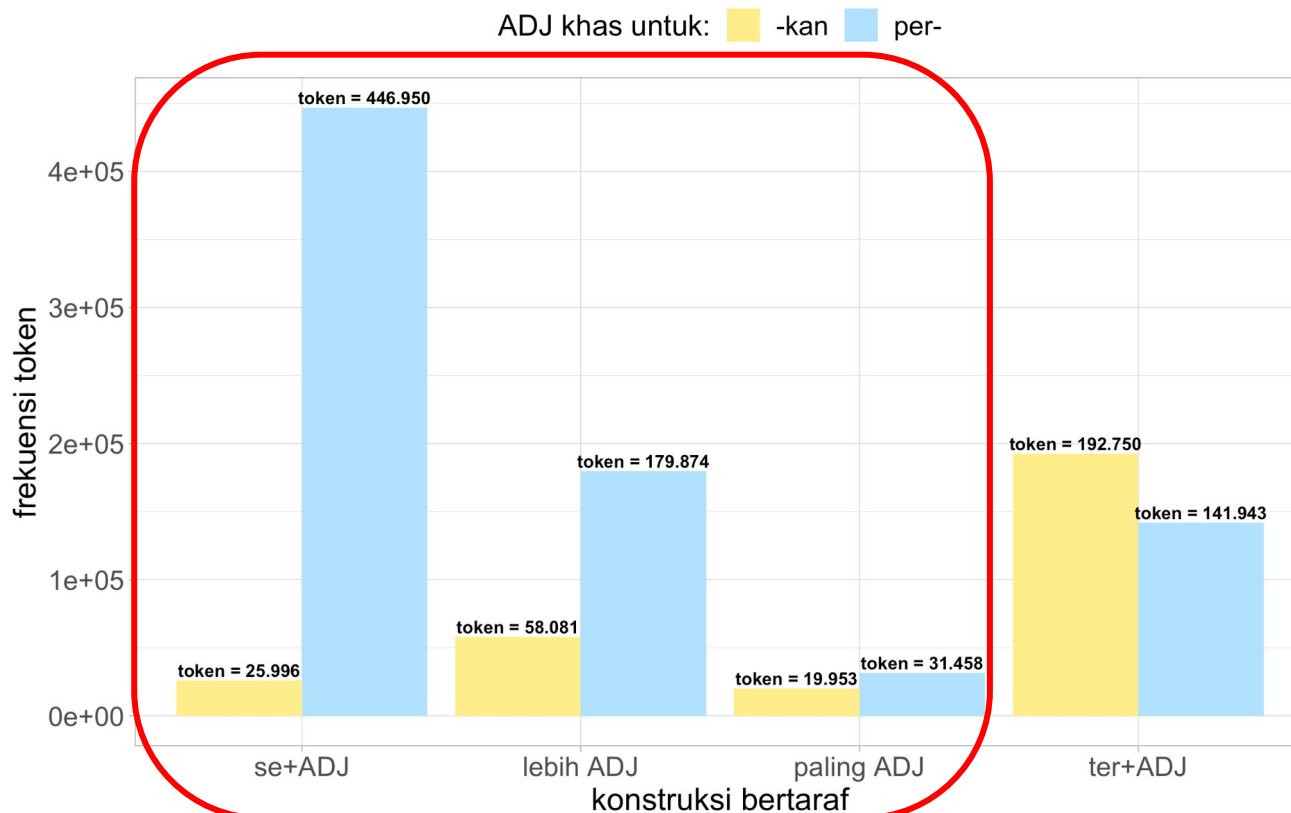
- Qualitative, and highly subjective, semantic intuition
- Other semantic types are present

Semantic characteristics of the distinctive adjectives for *per-* and *-kan*



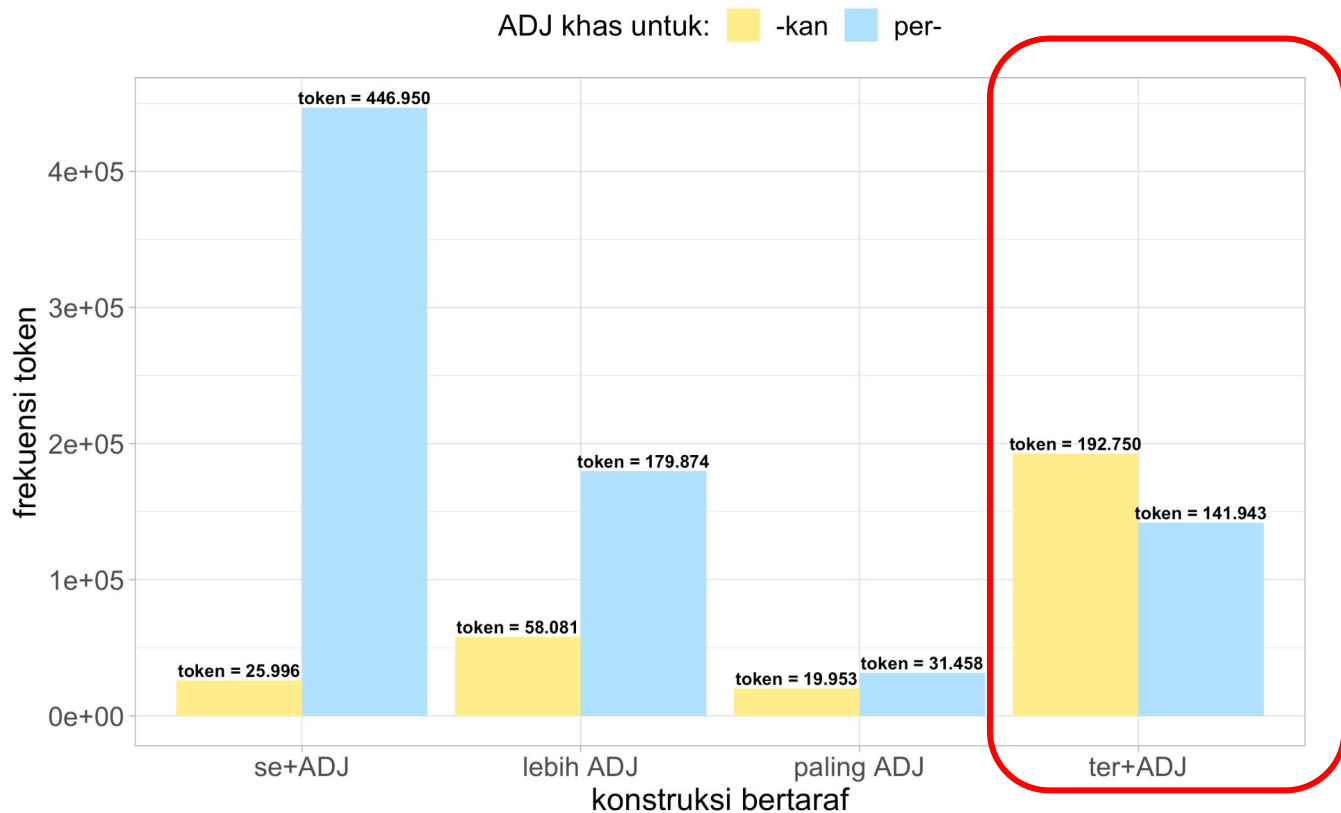
Perbedaan distribusi antara "-kan" & "per-" untuk tiap-tiap konstruksi sangat signifikan ($p < 0.001$; koreksi Holm)

Semantic characteristics of the distinctive adjectives for *per-* and *-kan*




Perbedaan distribusi antara "-kan" & "per-" untuk tiap-tiap konstruksi sangat signifikan ($p < 0.001$; koreksi Holm)

Semantic characteristics of the distinctive adjectives for *per-* and *-kan*



Perbedaan distribusi antara "-kan" & "per-" untuk tiap-tiap konstruksi sangat signifikan ($p < 0.001$; koreksi Holm)

- ~~1. Rival affixes~~
- ~~2. *per-* and *-kan*~~
- ~~3. Distinctive Collexeme Analysis~~
- ~~4. Results~~
5. Discussion & Conclusions 




Research questions

1. What are the adjectival bases that exhibit asymmetry for *per-* and *-kan*?

	ADJ khas untuk <i>-kan</i>	ADJ khas untuk <i>per-</i>
1	<i>nyata</i>	<i>kuat</i>
2	<i>jelas</i>	<i>cepat</i>
3	<i>tentu</i>	<i>luas</i>
4	<i>tetap</i>	<i>panjang</i>
5	<i>pasti</i>	<i>mudah</i>
6	<i>siap</i>	<i>banyak</i>
7	<i>umum</i>	<i>kaya</i>
8	<i>selamat</i>	<i>besar</i>
9	<i>sesuai</i>	<i>kecil</i>
10	<i>aman</i>	<i>ketat</i>

Research questions

2. What semantic preferences can be explored from the list of the distinctive adjectival bases for *per-* and *-kan*?

Adjective	<i>per-</i>	<i>-kan</i>
emotion/psych		
spatial / space / size / dimension		
gradability		

DISCUSSION

- *Distinctive Collexeme Analysis* is applicable to morphological phenomenon
 - Negative evidence for certain [affix+base word] configuration
 - Potential semantic preferences of *per-* and *-kan*
- Two rivals affixes could be distinctive quantitatively
 - Productivity (Denistia and Baayen, 2021; Rajeg & Denistia 2021)
 - Semantics (Rajeg & Rajeg 2019, Rajeg et al. 2019; Denistia et. al. 2021)
 - Functional load (Denistia and Baayen, 2021)
 - Distinctive bases
- Further study: experiment

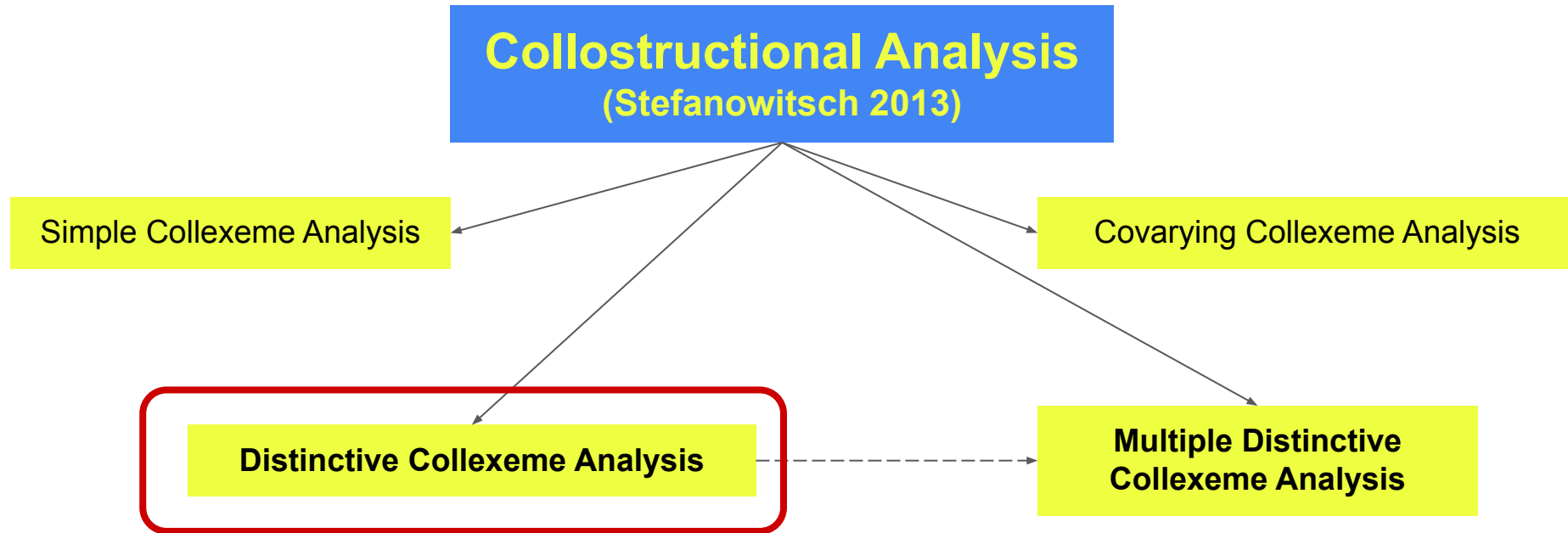
References

- Adelaar, K. A. (1992). *Proto malayic: The reconstruction of its phonology and parts of its lexicon and morphology*. Dept. of Linguistics, Research School of Pacific Studies, the Australian National University.
- Aronoff, M., & Anshen, F. (2017). Morphology and the Lexicon: Lexicalization and Productivity. In A. Spencer & A. M. Zwicky (Eds.), *The Handbook of Morphology* (pp. 237–247). Blackwell Publishing Ltd.
<https://doi.org/10.1002/9781405166348.ch11>
- Denistia, K., & Baayen, R. H. (2019). The Indonesian prefixes PE- and PEN-: A study in productivity and allomorphy. *Morphology*. <https://doi.org/10.1007/s11525-019-09340-7>
- Denistia, K., Shafaei-Bajestan, E., & Baayen, R. H. (2021). Exploring semantic differences between the Indonesian prefixes PE- and PEN- using a vector space model. *Corpus Linguistics and Linguistic Theory*.
<https://doi.org/10.1515/cllt-2020-0023>
- 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. Th., & Stefanowitsch, A. (2004). Extending collocation analysis: A corpus-based perspective on ‘alternations’. *International Journal of Corpus Linguistics*, 9(1), 97–129.
- Larasati, S. D., Kuboň, V., & Zeman, D. (2011). Indonesian Morphology Tool (MorphInd): Towards an Indonesian Corpus. *Systems and Frameworks for Computational Morphology*, 119–129. https://doi.org/10.1007/978-3-642-23138-4_8
- Nomoto, H., Choi, H., Moeljadi, D., & Bond, F. (2018). MALINDO Morph: Morphological dictionary and analyser for Malay/Indonesian. *Proceedings of the LREC 2018 Workshop ‘The 13th Workshop on Asian Language Resources’*, 36–43. http://lrec-conf.org/workshops/lrec2018/W29/pdf/8_W29.pdf

References

- Ogloblin, A. K. (1998). From Inert to Actional Causative. In L. Kulikov & H. Vater (Eds.), *Typology of Verbal Categories* (pp. 179–184). DE GRUYTER. <https://doi.org/10.1515/9783110913750.179>
- Rajeg, G. P. W., & Denistia, K. (2021, May 21). *A study in productivity of Indonesian causative per- and -kan* [Paper]. The 24th International Symposium on Malay/Indonesian Linguistics (ISMIL 24), Online. <https://doi.org/10.6084/m9.figshare.14633133>
- 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. *NUSA*, 67, 35–75.
- Rajeg, G. P. W., Denistia, K., & Rajeg, I. M. (2018). Working with a linguistic corpus using R: An introductory note with Indonesian negating construction. *Linguistik Indonesia*, 36(1), 1–36. <https://doi.org/10.26499/li.v36i1.71>
- Rajeg, G. P. W., & Rajeg, I. M. (2019). 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. (2021). Kajian linguistik korpus kuantitatif terhadap variasi ortografis dan kontekstual: Studi kasus untuk anggé/e(n) dan anggo(n) ‘pakai’ dalam Bahasa Bali. In I. N. Sudipa (Ed.), *Buku Persembahan untuk I Gede Sadia Menapaki Masa Purnabhakti* (pp. 366–388). Swasta Nulus. <https://doi.org/10.31219/osf.io/mu2fk>
- Sneddon, J. N., Adelaar, A., Djenar, D. N., & Ewing, M. C. (2010). *Indonesian reference grammar* (2nd ed.). Allen & Unwin.
- Stefanowitsch, A. (2006). Negative evidence and the raw frequency fallacy. *Corpus Linguistics and Linguistic Theory*, 2(1), 61–77.

Distinctive Collexeme Analysis (DCA) - *What is it?*



Distinctive Collexeme Analysis (DCA) - *How does it work?*

[*per*+ADJ] vs. [ADJ+*kan*]

Would these **observed distributions** (statistically) significantly differ from what would be expected IF there should be no distributional difference?

	<i>per</i> +ADJ	ADJ+ <i>kan</i>	Row Total	
<i>kaya</i>	2,628 (exp: 8.8% * 2,662)	34 (exp: 91.2% * 2,662)	2,662	Total <i>kaya</i>
Other adjectives	76,268 (exp: 8.8% * 896,604)	820,336 (exp: 91.2% * 896,604)	896,604	Total ¬ <i>kaya</i>
Column Total	78,896 8.8%	820,370 91.2%	899,266 100%	

Total freq. of
per-ADJ

Total freq. of
ADJ-*kan*

Combined freq.
of causative
morph. cxns

Distinctive Collexeme Analysis (DCA) - *How does it work?*

[*per*+**ADJ**] vs. [**ADJ**+*kan*]

Create 2-by-2 frequency table

```
> kaya <- c(per = 2628, kan = 34)
> other_adj <- c(per = 76268, kan = 820336)
> tbl <- rbind(kaya, other_adj)
> tbl
```

	per	kan
kaya	2628	34
other_adj	76268	820336

	<i>per-</i>	<i>-kan</i>
<i>kaya</i>	2,628	34
Other adjectives	76,268	820,336

Distinctive Collexeme Analysis (DCA) - *How does it work?*

[*per*+ADJ] vs. [ADJ+*kan*]

Compute expected frequency

```
> chisq.test(tbl)$expected
```

	per	kan
kaya	233.5473	2428.453
other_adj	78662.4527	817941.547

	<i>per-</i>	<i>-kan</i>
<i>kaya</i>	2,628 (exp: 234)	34 (exp: 2,428)
Other adjectives	76,268 (exp: 78,662)	820,336 (exp: 817,942)

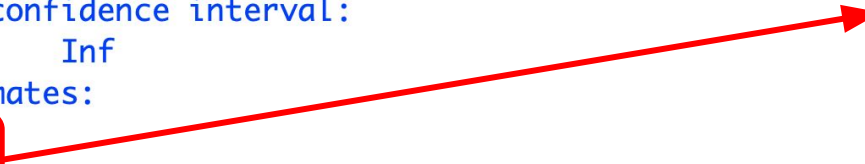
Distinctive Collexeme Analysis (DCA) - *How does it work?*

[*per*+**ADJ**] vs. [**ADJ**+*kan*]

```
> fisher.test(tbl, alternative = "greater")
```

Fisher's Exact Test for Count Data

```
data:  tbl
p-value < 2.2e-16
alternative hypothesis: true odds ratio is greater than 1
95 percent confidence interval:
 612.5714      Inf
sample estimates:
odds ratio
 829.8198
```



The likelihood to
observe PER-*kaya* is
829.8 times
GREATER than
kaya-KAN

```
> fisher.test(tbl, alternative = "greater")$p.value
[1] 0
> -log10(fisher.test(tbl, alternative = "greater")$p.value)
[1] Inf
```