

Dasar–dasar analisis korpus

Kata kunci, gugusan kata/n–grams, kolokasi

Gede Primahadi Wijaya RAJEG

Universitas Udayana, Bali

 <https://orcid.org/0000-0002-2047-8621>

 @PrimahadiWijaya

5 November 2020



Alur presentasi

- Daftar kata kunci
- Gugusan kata (*word cluster*) & n-grams
- Kolokasi & koligasi:
 - Contoh-contoh studi kasus pemanfaatan kolokasi



Kata kunci



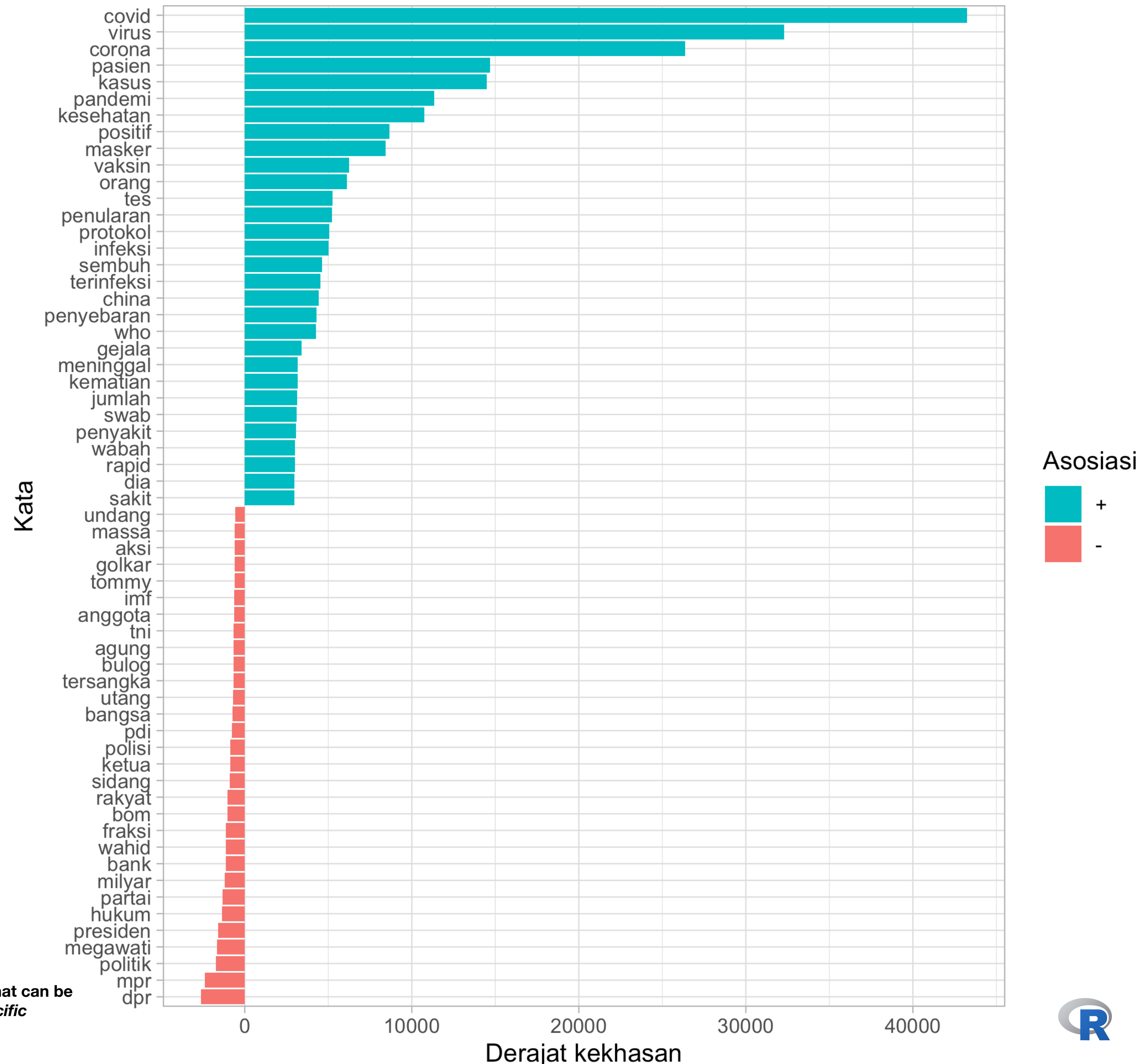
Kata kunci

- Bukanlah semata-mata kata-kata yang paling sering muncul di korpus
- Kata-kata yang “khas” di suatu korpus atas dasar penghitungan statistik
 - Perbandingan statistik daftar frekuensi kata di korpus target dengan korpus acuan (*reference/benchmark corpus*)
 - Log-likelihood atau Chi-Square
 - Kata kunci positif & kata kunci negatif
 - Kata kunci positif: signifikan secara statistik lebih sering muncul di korpus target
 - Kata kunci negatif: signifikan secara statistik lebih jarang muncul di korpus target

Kata kunci

- Kata kunci positif mencerminkan kosakata relevan untuk ranah/genre teks/korpus target
 - *Covid, virus, corona, pasien, kasus, pandemi, kesehatan, masker, vaksin, tes, penularan, protokol, infeksi, penyebaran, WHO, gejala, penyakit, wabah, ...*
 - Hasil yang tidak mengaggetkan
 - Korpus acuan: Kompas (umum) 2002-2003

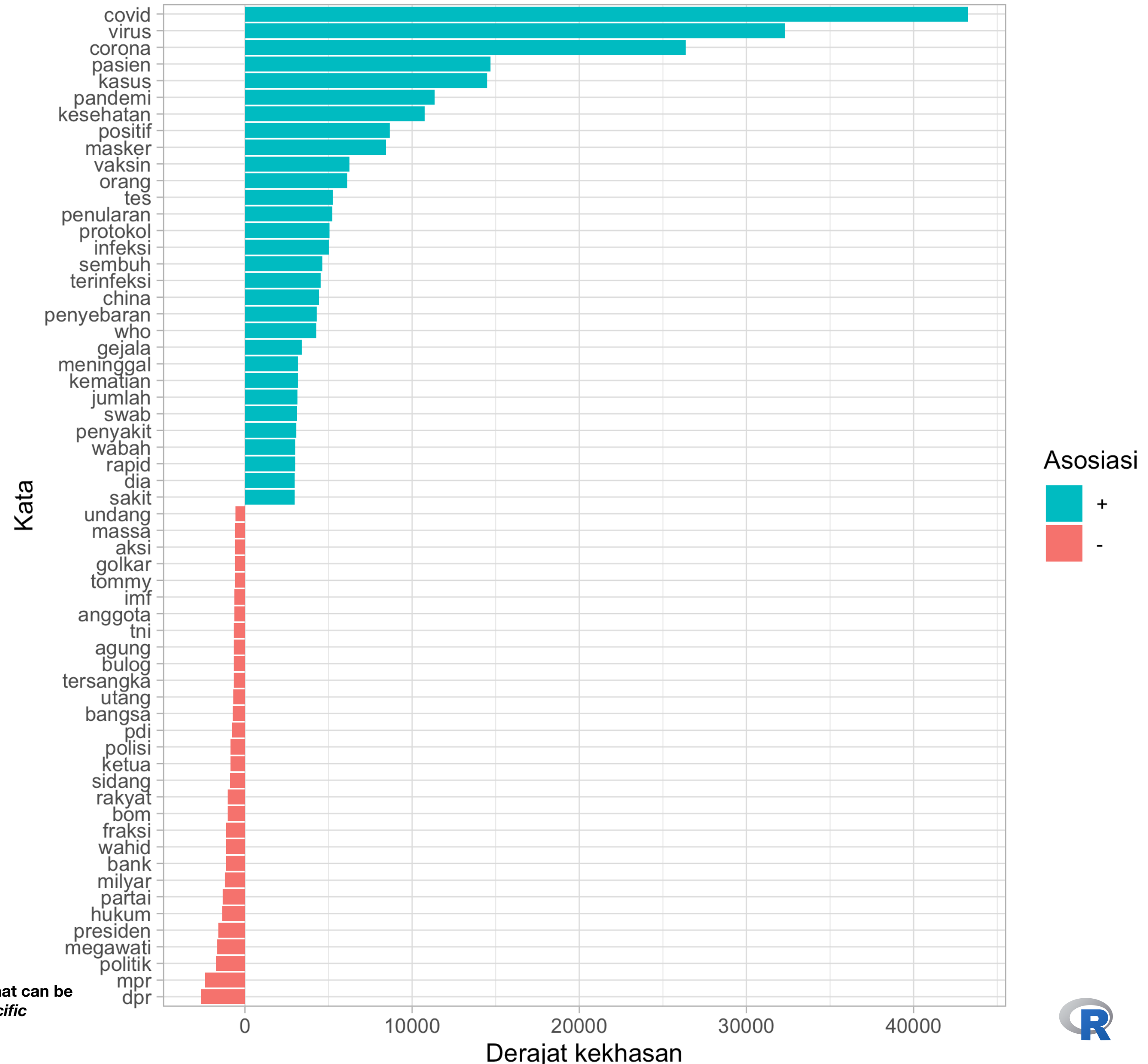
30 kata kunci positif dan negatif teratas di korpus Covid-19 (Arka, Sutarsa, & Rajeg 2020)



Kata kunci

- Kata kunci negatif mencerminkan kosakata yang TIDAK relevan untuk ranah/genre teks/korpus target
- Sebagian besar ranah **POLITIK, EKONOMI, HUKUM**

30 kata kunci positif dan negatif teratas di korpus Covid-19
(Arka, Sutarsa, & Rajeg 2020)





Gugusan kata & n-grams

Gugusan kata & n-grams

- Tataran analisis yang melewati kata tunggal

- Unit dua-kata (*bigram*)
- Unit tiga-kata (*trigram*)
- ...

merepresentasikan rangkaian kata yang cenderung muncul berangkaian (i.e. tanpa adanya unsur penyela di antara kata-kata tersebut)

- Tataran ini secara kolektif disebut dengan *n-grams*, *word clusters*, *lexical phrases* atau *lexical bundles*

N-grams

- N-gram bisa mencakup kolokasi

Mencetak gol/poin



- Namun n-gram bisa juga tidak merupakan kolokasi

! n-grams di dalam n-grams

babak perempat
babak perempat final



N-grams



- Perhatian terhadap batasan kalimat untuk n-gram yang bermakna secara sintaksis dan semantis
 - *There was a green house. The house was in a terrible state.*
 - *There was a green **house. The** house was in a terrible state.* (bi-gram)
 - *There was a green **house. The house** was in a terrible state.* (tri-gram)

Gugusan kata (*cluster*)

- Istilah yang lebih umum dan merangkul untuk dua hal berikut:
 - kolokasi n-gram (rangkaian kata yang bermakna)
 - rangkaian kata yang tidak begitu bermakna berdasarkan fungsi semantis dan sintaksis
 - *Lexical bundles*:
 - Konstruksi formulaik
 - *Would you like X; do you want X; can I have X*



Concordance

Concordance Plot

File View

Clusters/N-Grams

Collocates

Word List

Keyword List

Total No. of N-Gram Types 52

Total No. of N-Gram Tokens 231

Rank	Freq	N-gram
35	3	kebun pisang
36	3	lelaki yang
37	3	luh manik yang
38	3	manik yang
39	3	memutuskan untuk
40	3	para bule
41	3	petak sawah
42	3	rumah luh
43	3	satu satunya
44	3	seperti baling
45	3	seperti baling baling
46	3	takut kehilangan
47	3	tidak ada
48	3	tinggal di
49	3	untuk kemudian
50	3	warga desa
51	3	yang dulu
52	3	yang sedang

Cerpen Bunga Jepun (Kompas, 2003)

Search Term ☒ Words ☐ Case ☐ Regex ☒ N-Grams

Advanced

N-Gram Size

Min. 2 Max. 5

 Sort by ☐ Invert Order

Search Term Position

Sort by Freq

☒ On Left ☐ On Right

Min. Freq. Min. Range

3 1



Concordance

Concordance Plot

File View

Clusters/N-Grams

Collocates

Word List

Keyword List

Total No. of Cluster Types 38

Total No. of Cluster Tokens 328

Rank	Freq	Range	Prob	Cluster
1	94	1	0.072	dikenal dengan nama
2	29	1	0.022	dikenal dengan sebutan
3	18	1	0.014	dikenal dengan istilah
4	16	1	0.012	dikenal sebagai seorang
5	14	1	0.011	dikenal juga sebagai
6	12	1	0.009	dikenal sebagai salah
7	10	1	0.008	dikenal karena perannya
8	8	1	0.006	dikenal juga dengan
9	8	1	0.006	dikenal oleh masyarakat
10	8	1	0.006	dikenal sebagai orang
11	6	1	0.005	dikenal di kalangan
12	7	1	0.005	dikenal sebagai penulis
13	5	1	0.004	dikenal sebagai tokoh

**Cluster dengan kata target
dan fitur probabilitas transisi**

Search Term ☒ Words ☐ Case ☐ Regex ☐ N-Grams

dikenal

Advanced

Cluster Size

Min. 3

Max. 3

Start

Stop

Sort

Sort by ☐ Invert Order

Search Term Position

Sort by Prob

☒ On Left☐ On Right

Min. Freq. Min. Range

3

1



Alur presentasi

- Daftar kata kunci ✓
- Gugusan kata (*word cluster*) & n-grams ✓
- Kolokasi & koligasi:
 - Contoh-contoh studi kasus pemanfaatan kolokasi

Kolokasi (dan koligasi)

Kolokasi



shutterstock.com • 630652736

Fenomena bahwa kata-kata tertentu cenderung (*more likely*) muncul bersama dengan kata lainnya (Baker, Hardie, McEnery 2006: 36)



Kolokat



shutterstock.com • 630652736

Suatu kolokat adalah kata yang muncul di sekitaran (*within the neighbourhood*) kata lain (Baker, Hardie, McEnery 2006: 36-37).



Kolokasi



shutterstock.com • 630652736

Kemunculan bersama itu dilandasi atas adanya afiliasi/kedekatan di antara kedua kata tersebut.



Kolokasi sederhana



shutterstock.com • 630652736

- Afiliasi tersebut bisa mulai diamati dengan melihat kemunculan bersama kata-kata secara **berurutan** (*sequential co-occurrences of words*)
- Perbedaan secara semantis akan terlihat dari pilihan kombinasi kata berbeda terhadap kata target
- Nanti kita lihat definisi kuantitatif/statistik untuk **kolokasi**





Kolokasi sederhana

Kata
Dayak
Bangsa
Cadang
Bunga
Suku
Indian
Asli
Jawa
Melayu
Badui
Air



Concordance					Concordance Plot	File View	Clusters/N-Grams
Total No. of Cluster Types 24				Total No. of Cluster Tokens 247			
Rank	Freq	Range	Prob	Cluster			
1	54	1	0.112	suku kata			
2	24	1	0.050	suku-suku			
3	22	1	0.046	suku dayak			
4	17	1	0.035	suku bangsa			
5	14	1	0.029	suku ini			
6	13	1	0.027	suku cadang			
7	12	1	0.025	suku bunga			
8	10	1	0.021	suku di			
9	9	1	0.019	suku yang			
10	8	1	0.017	suku indian			
11	7	1	0.014	suku asli			
12	7	1	0.014	suku jawa			
13	7	1	0.014	suku melayu			
14	5	1	0.010	suku dan			
15	5	1	0.010	suku tersebut			
16	4	1	0.008	suku air			
17	4	1	0.008	suku badui			
18	4	1	0.008	suku dari			
19	4	1	0.008	suku hun			
20	4	1	0.008	suku kampai			

Search Term ☒ Words ☐ Case ☐ Regex ☐ N-Grams

suku

Advanced

Cluster Size

Min. 2 Max. 2

Min. Freq. 3 Min. Range 1

Sort by ☐ Invert Order

Search Term Position

Sort by Prob ☒ On Left ☐ On Right

Kolokat

Frekuensi kolokat suatu kata

- Dibatasi dalam rentang sisi kiri dan kanan kata target (*node*)
 - mis. Kata-kata yang muncul dalam rentang N-kata di kiri dan kanan kata target
- Dibatasi berdasarkan relasi sintaksis:
 - Kata target: verba
 - Kolokat: subjek, objek



shutterstock.com • 630652736



Kolokat

Frekuensi kolokat suatu kata



shutterstock.com • 630652736

	L4	L3	L2	L1	NODE	R1	R2	R3	R4	
di	mata	jaja	jelas	terpancar	kegembiraan	dan	kebahagiaan	bila	bersama	kau dan kami.

	L4	L3	L2	L1	NODE	R1	R2	
berteman	dengan	sesuatu	yang	membawa	kebahagiaan	semu	baginya	

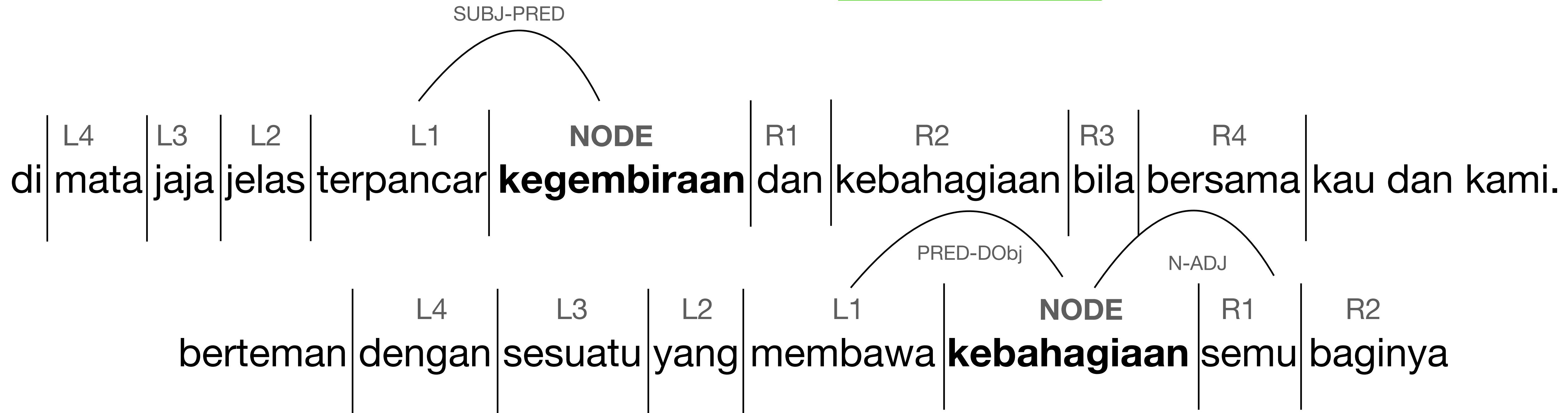


Kolokat

Frekuensi kolokat suatu kata



Kolokasi sintaksis



Kolokat

Frekuensi kolokat suatu kata



shutterstock.com • 630652736

A tibble: 48 x 9

w	l4	l3	l2	l1	r1	r2	r3	r4
<chr>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>
1 membawa	0	0	0	12	0	0	0	0
2 meraih	0	0	0	9	0	0	0	0
3 ia	0	0	8	0	0	0	0	0
4 dengan	0	0	0	0	0	0	0	6
5 saya	0	0	0	0	5	0	0	0
6 hidup	0	0	3	0	0	0	0	0
7 cita	0	0	0	0	0	0	3	0
8 kepuasan	0	0	0	0	0	0	3	0
9 paling	0	0	0	0	0	2	0	0
10 keberuntungan	0	0	2	0	0	0	0	0

... with 38 more rows



Kolokat rentang 4 kata di kiri dan kanan kata *kebahagiaan*

Koligasi





Koligasi



shutterstock.com • 630652736

Kemunculan bersama kata dengan kategori/fungsi gramatikal



Koligasi



shutterstock.com • 630652736

Tab. 43.6: Word-class colligates of *consequence* (incl. punctuation) in a random sample from the BNC World edition

LEFT COLLIGATES		RIGHT COLLIGATES	
Word class	F	Word class	F
Indefinite Article	46	<i>of</i>	55
Adjective	24	Punctuation Mark	26
Preposition	19	Personal Pronoun	4
Cardinal Numeral	5	Adverb	3
Definite Article	4	Preposition	2
<i>of</i>	2	<i>have</i> (3Sg)	2
		Indefinite Article	1
		Conjunction	1
		Possessive Pronoun	1
		Definite Article	1
		Noun (Singular)	1
		Quotation Mark	1
		Verb (3Sg)	1
		Verb (Past Tense)	1

Stefanowitsch & Gries (2009: 939)





Koligasi

Clusters Results 1				
Total No. of Cluster Types 29			Total No. of Cluster Tokens 411	
Rank	Freq	Range	Prob	Cluster
1	232	1	0.367	dianggap sebagai
2	32	1	0.051	dianggap tidak
3	15	1	0.024	dianggap telah
4	14	1	0.022	dianggap oleh
5	10	1	0.016	dianggap lebih
6	9	1	0.014	dianggap kurang
7	9	1	0.014	dianggap terlalu
8	7	1	0.011	dianggap merupakan
9	7	1	0.011	dianggap penting
10	6	1	0.009	dianggap menjadi
11	6	1	0.009	dianggap paling
12	5	1	0.008	dianggap cukup
13	5	1	0.008	dianggap dapat
14	5	1	0.008	dianggap sudah
15	4	1	0.006	dianggap berasal
16	4	1	0.006	dianggap gagal
17	4	1	0.006	dianggap sangat
18	4	1	0.006	dianggap sebuah
19	3	1	0.005	dianggap akan
20	3	1	0.005	dianggap bahwa
21	3	1	0.005	dianggap bentuk
22	3	1	0.005	dianggap identik
23	3	1	0.005	dianggap memenuhi
24	3	1	0.005	dianggap memiliki
25	3	1	0.005	dianggap mengganggu
26	3	1	0.005	dianggap satu
27	3	1	0.005	dianggap secara
28	3	1	0.005	dianggap suci
29	3	1	0.005	dianggap yang

Identifikasi dan pengelompokan (i.e. analisis) manual diperlukan

Mis.

dianggap bisa diikuti oleh Prep.
dianggap bisa diikuti oleh klausa

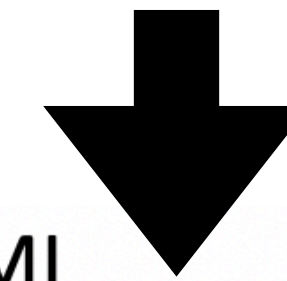
Kolokasi atas dasar analisis statistik

Pengukuran daya asosiasi keterkaitan kata

- Kemunculan bersama dua kata yang lebih sering dari yang diharapkan atas dasar kebetulan
- Melibatkan frekuensi kondisional:
 - Membandingkan frekuensi riil/pengamatan (*observed frequency*) dengan (ii) frekuensi harapan (*expected frequency*) atas dasar kebetulan
 - Menggunakan uji statistik untuk menentukan signifikansi derajat kedekatan antara kedua kata tersebut
 - Beberapa teknik pengukuran derajat kolokasi umum di Linguistik Korpus:
 - *Mutual information, t-score, z-score, log-likelihood*

Kolokasi atas dasar analisis statistik

Pengukuran daya asosiasi keterkaitan kata



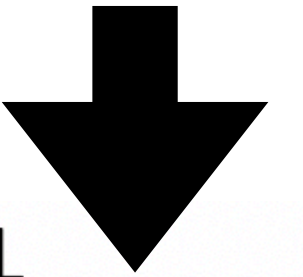
Word 1	Word 2	Freq Word 1 & Word 2	Freq Word 1	Freq Word 2	Word in Corpus	MI	T-Score	Z-Score	LL
tindak	pidana	702	1183	1251	5763049	11.417	26.486	1384.788	9019.440
terbukti	salah	67	773	568	5763049	9.780	8.176	242.462	761.350
tindak	kekerasan	198	1183	1573	5763049	9.260	14.048	347.877	2091.275
pihak	berwenang	289	6071	470	5763049	9.189	16.971	410.016	2942.094
jatuh	cinta	20	1061	325	5763049	8.385	4.459	81.518	191.025
melangkah	maju	9	141	1395	5763049	8.043	2.989	48.531	81.799
jatuh	hati	8	1061	1242	5763049	5.129	2.748	16.252	41.232

“Mutual information will give a high collocation score to relatively low-frequency word pairs” (Baker et al 2006: 38)



Kolokasi atas dasar analisis statistik

Pengukuran daya asosiasi keterkaitan kata



Word 1	Word 2	Freq Word 1 & Word 2	Freq Word 1	Freq Word 2	Word in Corpus	MI	T-Score	Z-Score	LL
tindak	pidana	702	1183	1251	5763049	11.417	26.486	1384.788	9019.440
pihak	berwenang	289	6071	470	5763049	9.189	16.971	410.016	2942.094
tindak	kekerasan	198	1183	1573	5763049	9.260	14.048	347.877	2091.275
terbukti	salah	67	773	568	5763049	9.780	8.176	242.462	761.350
jatuh	cinta	20	1061	325	5763049	8.385	4.459	81.518	191.025
melangkah	maju	9	141	1395	5763049	8.043	2.989	48.531	81.799
jatuh	hati	8	1061	1242	5763049	5.129	2.748	16.252	41.232

“whereas log-likelihood will give a higher score to higher-frequency pairs” (Baker et al 2006: 38)



Alur presentasi

- Daftar kata kunci ✓
- Gugusan kata (*word cluster*) & n-grams ✓
- Kolokasi & koligasi: ✓
 - Contoh-contoh studi kasus pemanfaatan data kolokasi

Tabel 5 Daftar 20 Kolokat Khas Teratas untuk *Kebahagiaan* di Korpus Leipzig Bahasa Indonesia

No.	Kolokat	N	exp	assocstr	p_binomial	p_holm	signif
1	<i>kesejahteraan</i>	82	32.701	29.646	2.258e-30	1.646e-25	***
2	<i>sejati</i>	89	37.703	26.595	2.538e-27	1.850e-22	***
3	<i>mencapai</i>	87	42.704	16.944	1.136e-17	8.282e-13	***
4	<i>akhirat</i>	66	29.624	16.829	1.482e-17	1.080e-12	***
5	<i>kesuksesan</i>	42	18.467	11.562	2.745e-12	2.000e-07	***
6	<i>kedamaian</i>	56	27.315	11.358	4.389e-12	3.198e-07	***
7	<i>menemukan</i>	68	35.394	11.260	5.496e-12	4.005e-07	***
8	<i>manusia</i>	87	50.399	10.034	9.257e-11	6.745e-06	***
9	<i>hidup</i>	172	116.956	9.889	1.290e-10	9.401e-06	***
10	<i>abadi</i>	49	24.238	9.626	2.365e-10	1.723e-05	***
11	<i>tangga</i>	37	16.928	9.191	6.447e-10	4.697e-05	***
12	<i>menuju</i>	45	22.314	8.829	1.483e-09	1.081e-04	***
13	<i>hakiki</i>	26	10.772	8.610	2.452e-09	1.786e-04	***
14	<i>dunia</i>	132	88.486	8.302	4.992e-09	3.637e-04	***
15	<i>keselamatan</i>	32	14.619	8.049	8.939e-09	6.512e-04	***
16	<i>kesehatan</i>	35	16.543	7.982	1.041e-08	7.584e-04	***
17	<i>meraih</i>	45	23.853	7.274	5.322e-08	3.876e-03	**
18	<i>umat</i>	26	11.542	7.151	7.063e-08	5.144e-03	**
19	<i>orang</i>	120	82.331	6.855	1.396e-07	1.016e-02	*
20	<i>merasakan</i>	92	60.017	6.752	1.769e-07	1.288e-02	*

Tabel 6 Daftar 20 Kolokat Khas Teratas untuk *kesenangan* di Korpus Leipzig Bahasa Indonesia

No.	Kolokat	N	exp	assocstr	p_binomial	p_holm	signif
1	<i>duniawi</i>	52	10.298	28.818	1.520e-29	1.108e-24	***
2	<i>pribadi</i>	30	5.994	16.702	1.985e-17	1.446e-12	***
3	<i>mencari</i>	55	17.522	15.757	1.748e-16	1.274e-11	***
4	<i>kenikmatan</i>	37	11.835	10.789	1.624e-11	1.184e-06	***
5	<i>hobi</i>	14	2.306	10.278	5.277e-11	3.845e-06	***

138

Linguistik Indonesia, Volume ke-38, No.2, Agustus 2020

No.	Kolokat	N	exp	assocstr	p_binomial	p_holm	signif
6	<i>nafsu</i>	19	3.996	10.114	7.689e-11	5.603e-06	***
7	<i>semata</i>	14	2.613	8.756	1.756e-09	1.279e-04	***
8	<i>keuntungan</i>	12	2.152	7.934	1.165e-08	8.488e-04	***
9	<i>kepentingan</i>	13	2.613	7.464	3.438e-08	2.504e-03	**
10	<i>seksual</i>	10	1.691	7.157	6.964e-08	5.072e-03	**
11	<i>menikmati</i>	42	18.291	7.136	7.312e-08	5.326e-03	**
12	<i>menunda</i>	8	1.230	6.507	3.115e-07	2.268e-02	*
13	<i>kebutuhan</i>	10	1.998	5.873	1.341e-06	9.762e-02	ms
14	<i>mengejar</i>	17	5.072	5.844	1.431e-06	1.042e-01	ns
15	<i>prinsip</i>	8	1.383	5.616	2.421e-06	1.762e-01	ns
16	<i>mendapat</i>	22	7.993	5.521	3.016e-06	2.196e-01	ns
17	<i>berdasarkan</i>	10	2.152	5.394	4.040e-06	2.941e-01	ns
18	<i>hawa</i>	9	1.844	5.171	6.744e-06	4.908e-01	ns
19	<i>waktu</i>	18	6.302	4.896	1.270e-05	9.242e-01	ns
20	<i>dosa</i>	6	0.922	4.880	1.319e-05	9.594e-01	ns



[*memperbesar/membesarkan* + word]

Tabel 4 Koleksem R1 khas signifikan untuk *membesarkan* dibandingkan dengan *memperbesar*

	koleksem	memperbesar	membesarkan	collstr
1	Partai	0 (41.53)	72 (30.47)	29.25
2	Hati	0 (19.03)	33 (13.97)	12.78
3	Anak	0 (10.38)	18 (7.62)	6.85
4	Pkb	0 (7.5)	13 (5.5)	4.92
5	namanya	0 (5.77)	10 (4.23)	3.77
6	Nama	0 (4.61)	8 (3.39)	3.01
7	anak-anak	0 (4.04)	7 (2.96)	2.63
8	Nu	0 (4.04)	7 (2.96)	2.63
9	Pan	1 (4.04)	6 (2.96)	1.60
10	anak-anaknya	0 (2.31)	4 (1.69)	1.50
11	Ketiga	0 (2.31)	4 (1.69)	1.50
12	organisasi	0 (2.31)	4 (1.69)	1.50
13	Pks	0 (2.31)	4 (1.69)	1.50

Tabel 5 Koleksem R1 khas signifikan untuk *memperbesar* dibandingkan dengan *membesarkan*

	W	memperbesar	membesarkan	collstr
1	keunggulan	107 (61.71)	0 (45.29)	28.47
2	kemenangan	46 (26.53)	0 (19.47)	11.48
3	Gol	23 (13.27)	0 (9.73)	5.61
4	Peluang	18 (10.38)	0 (7.62)	4.37
5	Pasar	8 (4.61)	0 (3.39)	1.92
6	Jumlah	9 (5.77)	1 (4.23)	1.48

[*memperbesar/membesarkan* + word]

Tabel 4 Koleksem R1 khas signifikan untuk *membesarkan* dibandingkan dengan *memperbesar*

	koleksem	memperbesar	membesarkan	collstr
1	Partai	0 (41.53)	72 (30.47)	29.25
2	Hati	0 (19.03)	33 (13.97)	12.78
3	Anak	0 (10.38)	18 (7.62)	6.85
4	Pkb	0 (7.5)	13 (5.5)	4.92
5	namanya	0 (5.77)	10 (4.23)	3.77
6	Nama	0 (4.61)	8 (3.39)	3.01
7	anak-anak	0 (4.04)	7 (2.96)	2.63
8	Nu	0 (4.04)	7 (2.96)	2.63
9	Pan	1 (4.04)	6 (2.96)	1.60
10	anak-anaknya	0 (2.31)	4 (1.69)	1.50
11	Ketiga	0 (2.31)	4 (1.69)	1.50
12	organisasi	0 (2.31)	4 (1.69)	1.50
13	Pks	0 (2.31)	4 (1.69)	1.50

IMPORTANCE IS BIG (Lakoff & Johnson, 1999, p. 50; Lakoff & Núñez, 2000, p. 41)

Tabel 5 Koleksem R1 khas signifikan untuk *memperbesar* dibandingkan dengan *membesarkan*

	W	memperbesar	membesarkan	collstr
1	keunggulan	107 (61.71)	0 (45.29)	28.47
2	kemenangan	46 (26.53)	0 (19.47)	11.48
3	Gol	23 (13.27)	0 (9.73)	5.61
4	Peluang	18 (10.38)	0 (7.62)	4.37
5	Pasar	8 (4.61)	0 (3.39)	1.92
6	Jumlah	9 (5.77)	1 (4.23)	1.48

QUANTITY IS SIZE
↓
MORE IS BIG

(Winter & Matlock, 2013; Lakoff & Núñez, 2000, pp. 55-56)

- Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. Basic Books.
- Lakoff, G., & Núñez, R. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being*. Basic Books.
- Winter, B., & Matlock, T. (2013). Creativity and the sensorimotor grounding of mathematics. In M. Borkent, B. Dancygier, & J. Hinnell (Eds.), *Language and the creative mind* (pp. 37–48). CSLI Publications. http://www.bodowinter.com/papers/bwtm_creativity_math.pdf

[*memperbesar/membesarkan* + word]

Tabel 4 Koleksem R1 khas signifikan untuk *membesarkan* dibandingkan dengan *memperbesar*

	koleksem	<i>memperbesar</i>	<i>membesarkan</i>	colloc
1	Partai	0	0	0
2	Hati	0	0	0
3	Anak	0	0	0
4	Pkb	0	0	0
5	namanya	0	0	0
6	Nama	0	0	0
7	anak-anak	0	0	0
8	Nu	0	0	0
9	Pan	1 (4.04)	6 (2.96)	1.60
10	anak-anaknya	0 (2.31)	4 (1.69)	1.50
11	Ketiga	0 (2.31)	4 (1.69)	1.50
12	organisasi	0 (2.31)	4 (1.69)	1.50
13	Pks	0 (2.31)	4 (1.69)	1.50

Tabel 5 Koleksem R1 khas signifikan untuk *memperbesar* dibandingkan dengan *membesarkan*

	koleksem	<i>memperbesar</i>	<i>membesarkan</i>	colloc
1	Partai	0	0	0
2	Hati	0	0	0
3	Anak	0	0	0
4	Pkb	0	0	0
5	namanya	0	0	0
6	Nama	0	0	0
7	anak-anak	0	0	0
8	Nu	0	0	0
9	Pan	1 (4.04)	6 (2.96)	1.60
10	anak-anaknya	0 (2.31)	4 (1.69)	1.50
11	Ketiga	0 (2.31)	4 (1.69)	1.50
12	organisasi	0 (2.31)	4 (1.69)	1.50
13	Pks	0 (2.31)	4 (1.69)	1.50

```
> filter(colloc_out, w %in% "ketiga") %>% pull(sent_match)
```

[1] "46142 Tiga tahun sejak Septi dilahirkan, Fatma terpaksa **membesarkan ketiga anaknya** setelah ditinggal cerai suami nya yang malu karena memiliki anak dengan fisik yang berbeda dengan anak normal lainnya."

[2] "126100 Wanita yang tetap terlihat cantik di usianya yang 40 tahun itu kemudian mengaku bahwa ia untuk sementara memilih untuk mendukung suaminya saja sambil **membesarkan ketiga anaknya** yang masih relatif kecil-kecil."

[3] "297342 Suami Victoria Beckham, penyanyi dan model kondang asal Inggris, itu mengisyaratkan akan tinggal lebih lama di LA karena tempat itu merupakan tempat yang nyaman untuk **membesarkan ketiga puteranya.**"

[4] "151020 Bintang olah raga Inggris berusia 35 tahun dan istrinya yang berusia 36 tahun serta telah beralih dari penyanyi ke perancang busana itu kini tengah **membesarkan ketiga anak laki-laknya**, yaitu Brooklyn (12), Romeo (8), dan Cruz (6)."

Nama *besar*; hari *raya*; hari *besar* keagamaan; rakyat *kecil*

IMPORTANCE IS BIG (Lakoff & Johnson, 1999, p. 50; Lakoff & Núñez, 2000, p. 41)

- Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. Basic Books.
- Lakoff, G., & Núñez, R. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being*. Basic Books.

[*memperbesar/membesarkan* + word]

Tabel 4 Koleksem R1 khas signifikan untuk *membesarkan* dibandingkan dengan *memperbesar*

	koleksem	memperbesar	membesarkan	collstr
1	Partai	0 (41.53)	72 (30.47)	29.25
2	Hati	0 (19.03)	33 (13.97)	12.78
3	Anak	0 (10.38)	18 (7.62)	6.85
4	Pkb	0 (7.5)	13 (5.5)	4.92
5	namanya	0 (5.77)	10 (4.23)	3.77
6	Nama	0 (4.61)	8 (3.39)	3.01
7	anak-anak	0 (4.04)	7 (2.96)	2.63
8	Nu	0 (4.04)	7 (2.96)	2.63
9	Pan	1 (4.04)	6 (2.96)	1.60
10	anak-anaknya	0 (2.31)	4 (1.69)	1.50
11	Ketiga	0 (2.31)	4 (1.69)	1.50
12	organisasi	0 (2.31)	4 (1.69)	1.50
13	Pks	0 (2.31)	4 (1.69)	1.50

Tabel 5 Koleksem R1 khas signifikan untuk *memperbesar* dibandingkan dengan *membesarkan*

	W	memperbesar	membesarkan	collstr
1	keunggulan	107 (61.71)	0 (45.29)	28.47
2	kemenangan	46 (26.53)	0 (19.47)	11.48
3	Gol	23 (13.27)	0 (9.73)	5.61
4	Peluang	18 (10.38)	0 (7.62)	4.37
5	Pasar	8 (4.61)	0 (3.39)	1.92
6	Jumlah	9 (5.77)	1 (4.23)	1.48

QUANTITY IS SIZE
↓
MORE IS BIG

(Winter & Matlock, 2013; Lakoff & Núñez, 2000, pp. 55-56)

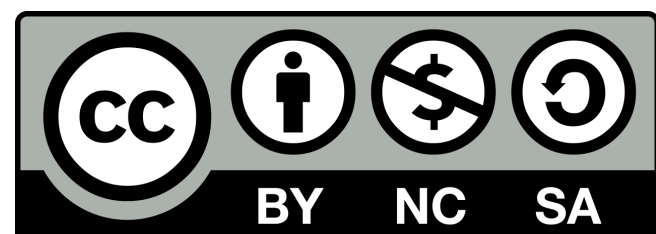
diskon *besar-besaran*; makan *besar*

- Lakoff, G., & Núñez, R. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being*. Basic Books.
- Winter, B., & Matlock, T. (2013). Creativity and the sensorimotor grounding of mathematics. In M. Borkent, B. Dancygier, & J. Hinnell (Eds.), *Language and the creative mind* (pp. 37–48). CSLI Publications. http://www.bodowinter.com/papers/bwtm_creativity_math.pdf



Alur presentasi

- Daftar kata kunci ✓
- Gugusan kata (*word cluster*) & n-grams ✓
- Kolokasi & koligasi: ✓
 - Contoh-contoh studi kasus pemanfaatan data kolokasi ✓



Dasar–dasar analisis korpus

Kata kunci, gugusan kata/n-grams, kolokasi

Gede Primahadi Wijaya RAJEG

Universitas Udayana, Bali

 <https://orcid.org/0000-0002-2047-8621>

 @PrimahadiWijaya

5 November 2020