
UNIVERSITY
OF OULU


IARIA

! " # $
% & '

& (

%

" #
)& #"*" ++ +,- ,,./+0.12. ,-3142-
* #

5 6#$5 & 7 ! 6# 8" # 96 # 8
" # :
;<6 % * &

$ & 7'

" %) %) %

; & % # %

& % = # "

!"

									8
						8		8	
8	8	!	8						
			8					8	8
!									
								!	

#

$

7 &"% &"
    % >   % ;
)"
 &
= 7 8

      & (?

      7 ! &   0-,/
)@

" %&' "

$ & '; ? * A

) ,0

= % 2 B

,0 8

( )#

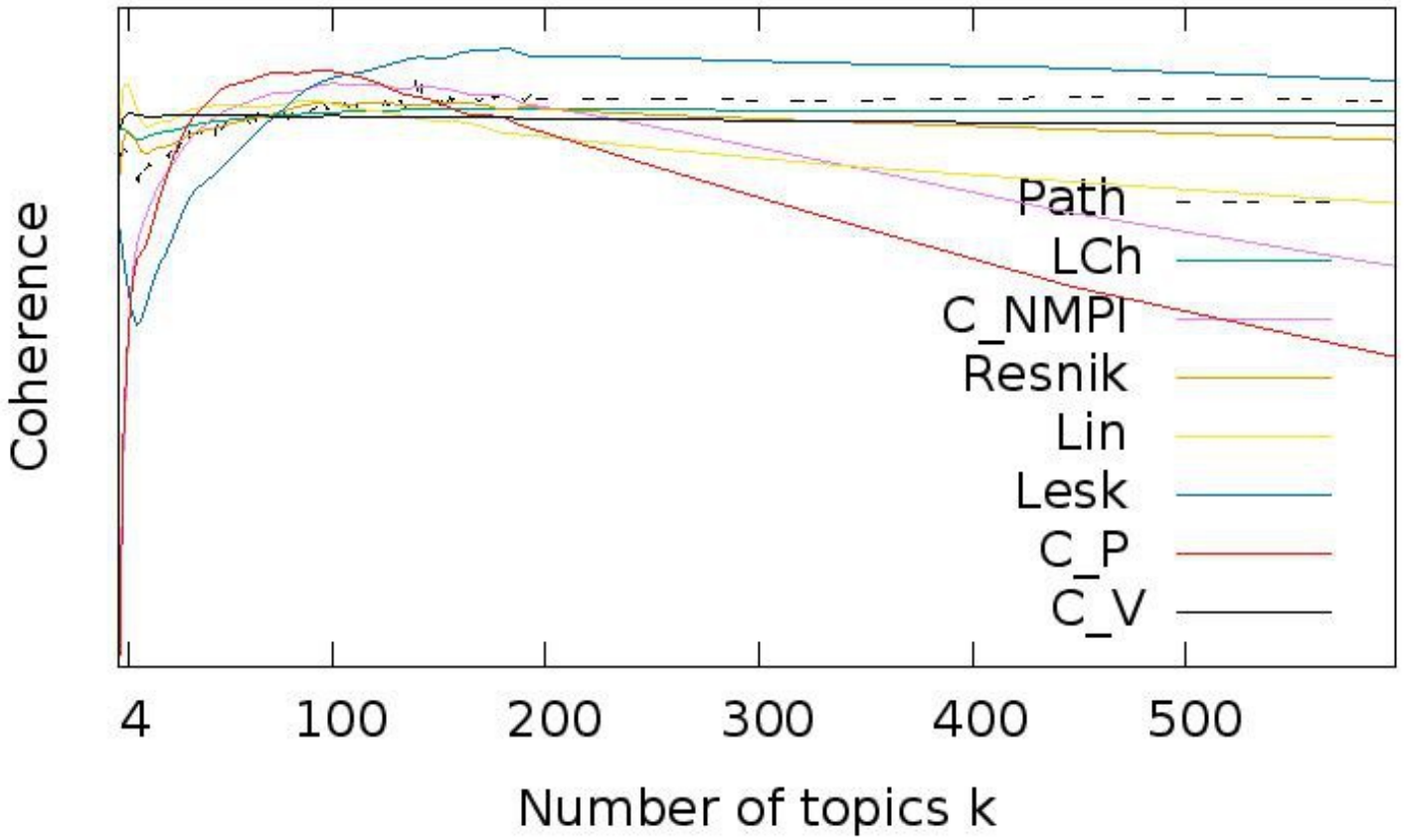

Coherence
Path
LCh
C_NMPI
Resnik
Lin
Lesk
C_P
C_V
4
100
200
300
400
500
Number of topics k

TABLE I. k-VALUES OF THREE HIGHEST COHERENCE VALUES FOR 12 CORPORA (A - D10) GIVEN BY 16 COHERENCE MEASURES ($HsO - C_{UMass}$).

	HsO	LCh	Lesk	WuP	Resnik	JCn	Lin	Path	vec_p	vec	C_A	C_P	C_V	C_{NPMI}	C_{UCI}	C_{UMass}
A	95	179	112	23	23	7	6	179	116	116	14	93	7	172	172	6
	112	450	115	7	6	23	7	450	178	102	9	138	9	95	181	25
	102	139	174	6	7	6	23	139	144	108	12	95	21	181	162	23
A20	164	146	143	7	7	6	7	143	124	36	12	99	4	95	95	77
	19	143	164	8	6	59	6	146	144	87	10	64	16	99	77	64
	90	132	173	10	12	7	93	144	146	60	7	151	24	59	183	62
A10	175	187	93	37	37	8	37	187	129	4	20	51	6	51	120	32
	93	145	164	93	93	93	8	145	163	6	51	102	76	88	51	52
	37	175	137	175	112	4	93	175	4	129	24	114	80	120	88	61
B	150	198	198	108	62	58	90	198	250	85	69	69	5	69	69	11
	196	147	164	89	90	54	62	147	92	89	7	11	6	81	158	10
	117	161	196	109	89	52	89	161	280	135	6	46	22	158	11	26
B20	170	143	129	33	33	33	33	149	5	67	10	101	5	66	101	10
	171	149	149	8	109	25	109	143	153	60	4	66	12	101	102	9
	190	187	171	5	63	48	5	170	181	142	14	10	6	95	95	25
B10	73	127	127	73	73	116	31	127	4	4	73	11	14	80	80	10
	146	146	181	64	105	31	73	147	5	135	23	80	7	88	68	11
	175	175	164	105	175	21	105	146	135	6	48	10	20	68	88	12
C	140	7	140	7	140	32	70	7	144	133	9	68	5	107	107	26
	155	8	193	70	113	104	98	5	143	106	17	107	11	140	140	41
	113	5	192	140	70	80	140	8	160	132	12	40	28	126	126	35
C20	50	153	188	11	11	11	11	133	132	111	8	140	5	157	157	33
	157	133	180	50	48	50	50	166	86	67	13	67	9	140	96	8
	144	166	144	48	50	10	48	153	133	152	14	81	7	96	140	14
C10	66	164	66	6	12	121	6	157	64	37	21	42	4	21	21	29
	69	189	103	17	140	4	12	189	117	48	9	9	8	16	22	19
	90	145	185	152	16	28	89	164	25	99	8	112	5	19	69	74
D	100	166	188	6	6	6	6	166	135	83	113	103	12	92	92	17
	149	7	191	7	10	183	7	143	185	146	9	113	113	113	189	19
	188	6	100	9	7	54	8	188	198	167	8	32	103	162	196	24
D20	97	116	107	48	48	4	48	144	7	72	12	78	41	109	109	14
	118	144	144	73	73	48	73	116	29	106	4	73	39	73	73	7
	107	169	184	20	91	91	46	169	197	91	16	55	40	100	100	16
D10	90	69	77	22	77	32	36	69	26	15	12	69	7	57	58	5
	79	141	79	36	90	36	12	280	39	45	6	57	8	58	57	64
	93	67	69	12	67	29	22	141	41	57	18	58	47	69	20	67

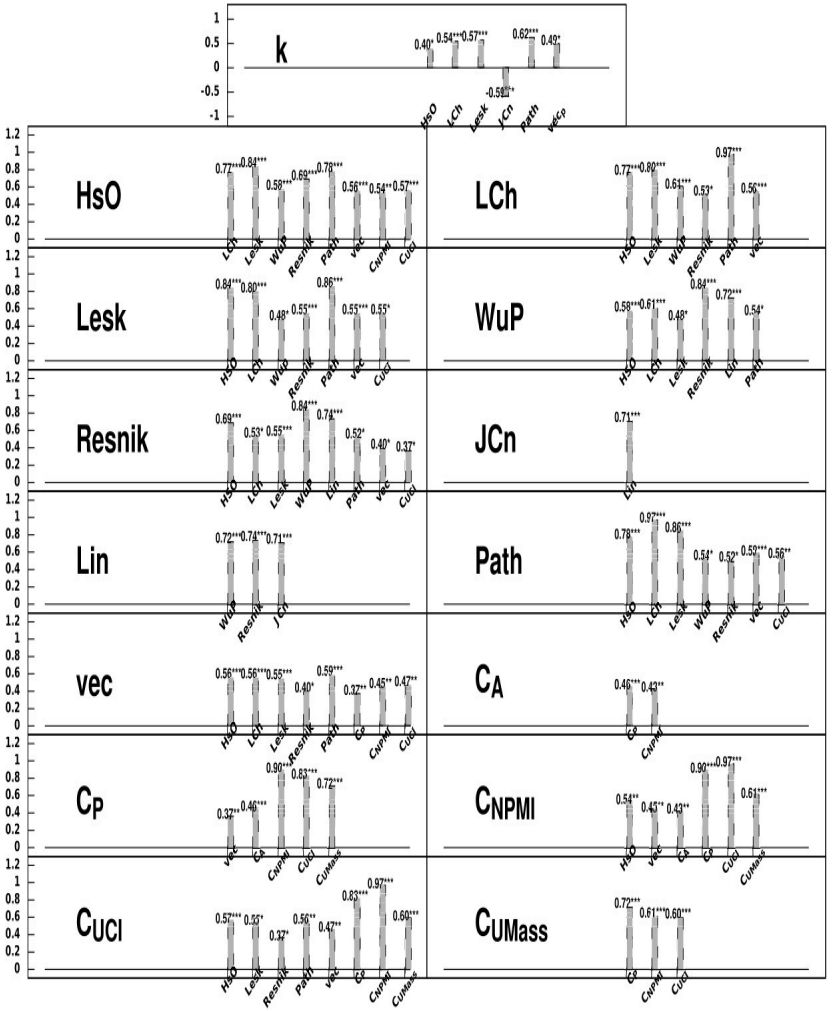

k
HsO
LCh
Lesk
WuP
Resnik
JCn
Lin
Path
vec
CA
CP
CNPMI
CUCI
CUMass

TABLE V. PEARSON AND SPEARMAN CORRELATIONS BETWEEN FOUR HUMAN RATINGS (MC - SIMLEX NOUNS) AND 16 COHERENCE MEASURES ($HsO - C_{\text{UMASS}}$). NOTE: HERE VALUES **without any** ASTERISKS ARE STATISTICALLY HIGHLY SIGNIFICANT WITH P<0.001. AND ** : P<0.01, AND * : P<0.05 , – : P>0.05 AND N.D. MEANS NO DATA.

	HsO	LCh	Lesk	WuP	Resnik	JCn	Lin	Path	vec_p	vec	C_{A}	C_{P}	C_{V}	C_{NPMI}	C_{UCI}	C_{UMass}
MC(P)	–	0.57*	–	0.55*	0.59	–	0.53*	–	0.60	**0.88**	–	0.79	–	0.77	0.67	–
MC(S)	–	0.58*	0.60	0.55*	0.68	–	0.56*	0.56*	0.70	**0.90**	–	0.81	0.65	0.82	–	–
RG(P)	0.54	0.60	0.44	0.53	0.61	–	0.54	0.54	n.d.	n.d.	–	0.75	–	**0.77**	0.71	–
RG(S)	0.49	0.56	0.55	0.51	0.55	–	0.46	0.54	n.d.	n.d.	–	**0.85**	0.50	0.84	0.83	0.45
Lau(P)	0.19	–	0.15	0.18	0.25	0.33	0.29	–	n.d.	n.d	0.38	**0.61**	0.31	0.55	0.51	0.28
Lau(S)	0.25	–	0.19	0.20	0.31	0.39	0.37	–	n.d.	n.d.	0.39	**0.52**	0.33	0.49	0.46	0.26
Simlex n.(P)	0.35	**0.52**	0.25	0.45	0.41	0.35	0.51	0.51	0.28	0.35	–	0.24	0.13	0.17	0.18	–
Simlex n.(S)	0.36	0.49	0.31	0.47	0.41	**0.51**	**0.51**	0.48	0.22	0.33	–	0.22	0.21	0.16	0.18	–

$

%

= 7 8 8 8

' % B ?

C B

4--- + 0 8

!

D,--% ( )@

E ! 8

,1

! < $! (

; 8


UNIVERSITY
OF OULU


IARIA

! "

#

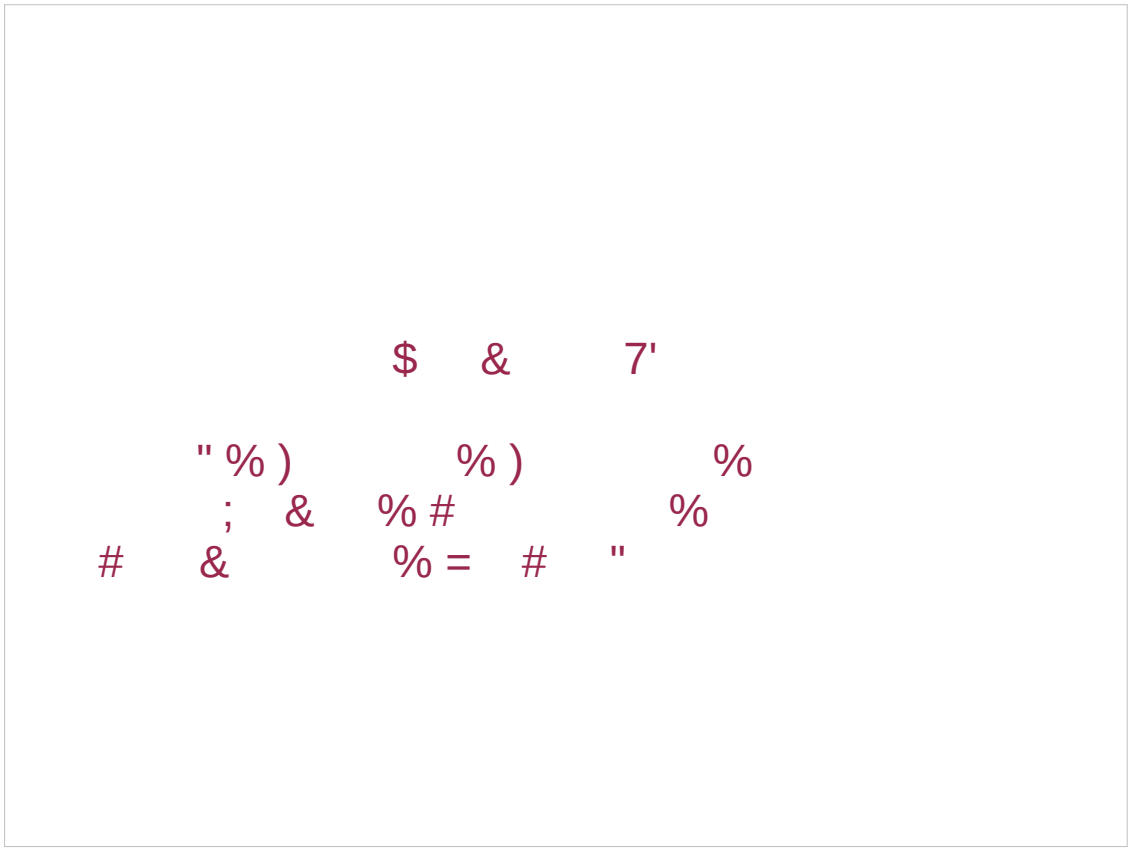
! " # $
% & '

& (

%

" #
)& #"*" ++ +,- ,,./+0.12. ,-3142-
* #

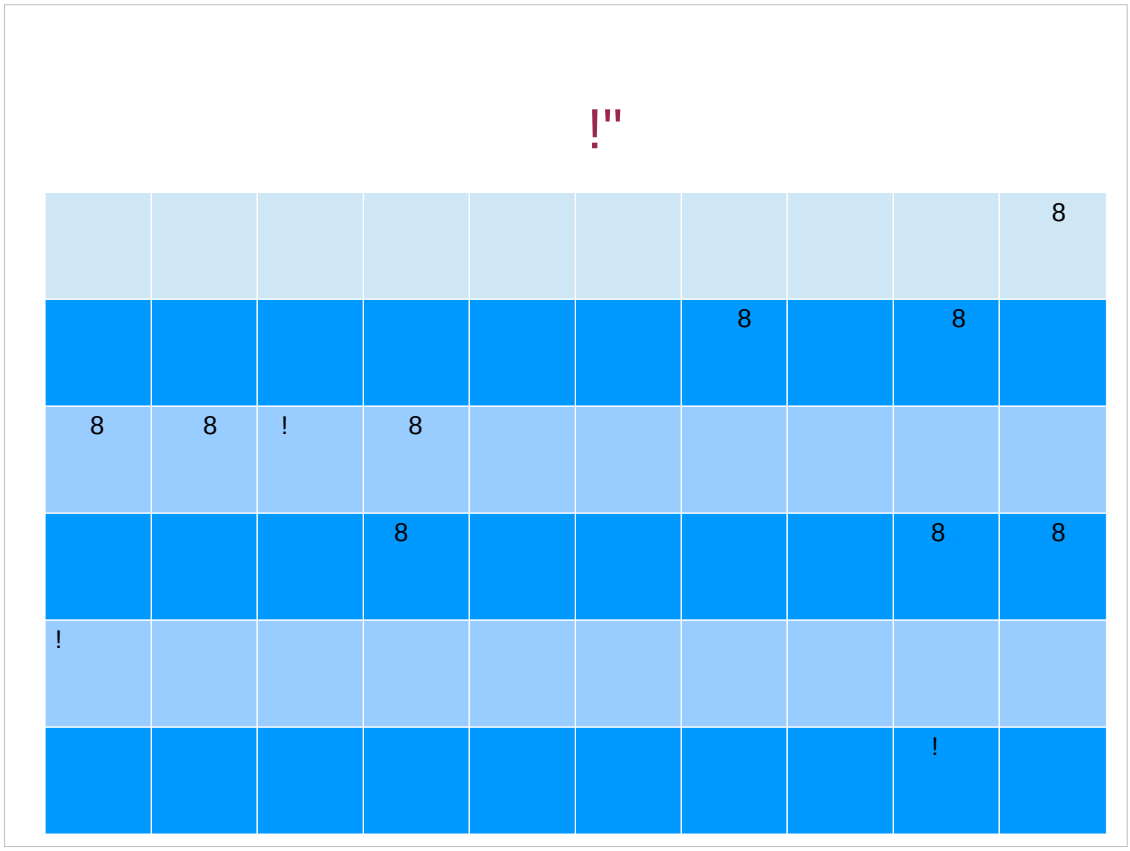
5 6#$5 & 7 ! 6# 8" # 96 # 8
" # :
;<6 % * &


$ & 7'
" %) %) %
; & % # %
& % = # "

$
$

! "

#


!"
8
8
8
8
8
!
8
8
8
8
!
!

% & '('&
'(

% ")

%

#

$

7 &"% &"
% > % ;
)"
&
= 7 8

& (?

7 ! & 0-,/
)@

*

"

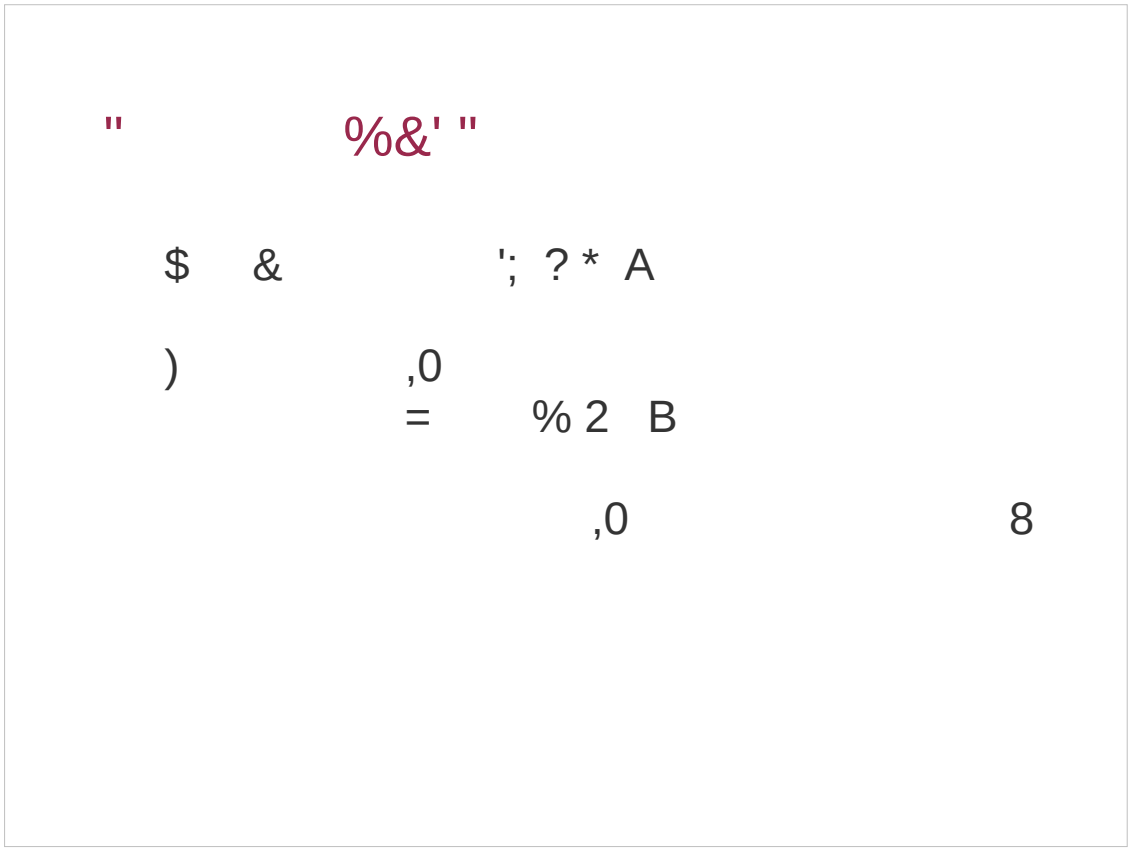
%

')
'(+ $) ,

*

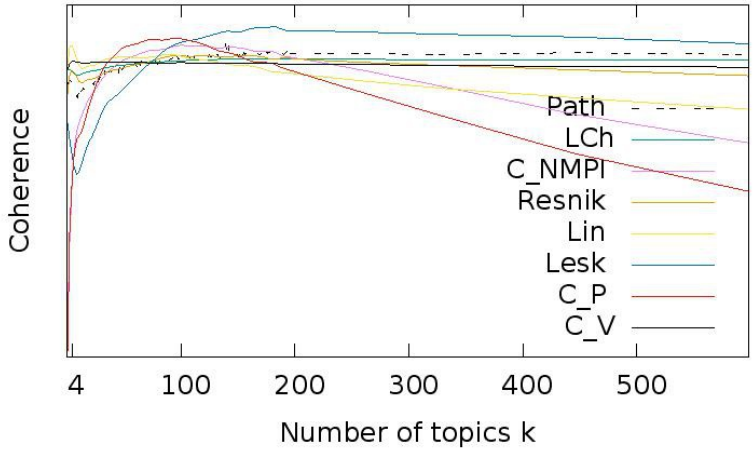
, -.

,



* "
/ 0 % 1/0%2 3 4
- '5 "
6 #
7 '5

( )#


Coherence
Path
LCh
C_NMPI
Resnik
Lin
Lesk
C_P
C_V
4
100
200
300
400
500
Number of topics k

8 " 9'((

% " -. " 9:(

, /-

"

TABLE I. k-VALUES OF THREE HIGHEST COHERENCE VALUES FOR 12 CORPORA (A - D10) GIVEN BY 16 COHERENCE MEASURES ($HsO - C_{UMass}$).

	HsO	LCh	Lesk	WuP	Resnik	JCn	Lin	Path	vec_p	vec	C_A	C_P	C_V	C_{NPMI}	C_{UCI}	C_{UMass}
A	95	179	112	23	23	7	6	179	116	116	14	93	7	172	172	6
	112	450	115	7	6	23	7	450	178	102	9	138	9	95	181	25
	102	139	174	6	7	6	23	139	144	108	12	95	21	181	162	23
A20	164	146	143	7	7	6	7	143	124	36	12	99	4	95	95	77
	19	143	164	8	6	59	6	146	144	87	10	64	16	99	77	64
	90	132	173	10	12	7	93	144	146	60	7	151	24	59	183	62
A10	175	187	93	37	37	8	37	187	129	4	20	51	6	51	120	32
	93	145	164	93	93	93	8	145	163	6	51	102	76	88	51	52
	37	175	137	175	112	4	93	175	4	129	24	114	80	120	88	61
B	150	198	198	108	62	58	90	198	250	85	69	69	5	69	69	11
	196	147	164	89	90	54	62	147	92	89	7	11	6	81	158	10
	117	161	196	109	89	52	89	161	280	135	6	46	22	158	11	26
B20	170	143	129	33	33	33	33	149	5	67	10	101	5	66	101	10
	171	149	149	8	109	25	109	143	153	60	4	66	12	101	102	9
	190	187	171	5	63	48	5	170	181	142	14	10	6	95	95	25
B10	73	127	127	73	73	116	31	127	4	4	73	11	14	80	80	10
	146	146	181	64	105	31	73	147	5	135	23	80	7	88	68	11
	175	175	164	105	175	21	105	146	135	6	48	10	20	68	88	12
C	140	7	140	7	140	32	70	7	144	133	9	68	5	107	107	26
	155	8	193	70	113	104	98	5	143	106	17	107	11	140	140	41
	113	5	192	140	70	80	140	8	160	132	12	40	28	126	126	35
C20	50	153	188	11	11	11	11	133	132	111	8	140	5	157	157	33
	157	133	180	50	48	50	50	166	86	67	13	67	9	140	96	8
	144	166	144	48	50	10	48	153	133	152	14	81	7	96	140	14
C10	66	164	66	6	12	121	6	157	64	37	21	42	4	21	21	29
	69	189	103	17	140	4	12	189	117	48	9	9	8	16	22	19
	90	145	185	152	16	28	89	164	25	99	8	112	5	19	69	74
D	100	166	188	6	6	6	6	166	135	83	113	103	12	92	92	17
	149	7	191	7	10	183	7	143	185	146	9	113	113	113	189	19
	188	6	100	9	7	54	8	188	198	167	8	32	103	162	196	24
D20	97	116	107	48	48	4	48	144	7	72	12	78	41	109	109	14
	118	144	144	73	73	48	73	116	29	106	4	73	39	73	73	7
	107	169	184	20	91	91	46	169	197	91	16	55	40	100	100	16
D10	90	69	77	22	77	32	36	69	26	15	12	69	7	57	58	5
	79	141	79	36	90	36	12	280	39	45	6	57	8	58	57	64
	93	67	69	12	67	29	22	141	41	57	18	58	47	69	20	67

"

" $ 6 $

'5

- $

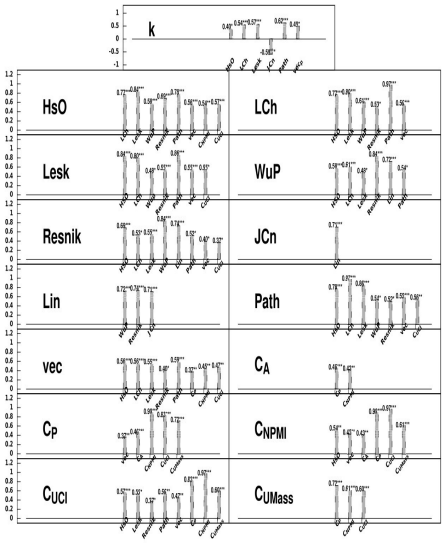
1 + $, 2

$ $

/- ,

-. .

0 7$!!


k
HsO
LCh
Lesk
WuP
Resnik
JCn
Lin
Path
vec
C_A
C_P
C_{NPMI}
C_{UCI}
C_{UMass}

0 7$ $
!!

+ $

, $ -.

7

"

* $

TABLE V. PEARSON AND SPEARMAN CORRELATIONS BETWEEN FOUR HUMAN RATINGS (MC - SIMLEX NOUNS) AND 16 COHERENCE MEASURES ($HsO - C_{UMASS}$). NOTE: HERE VALUES **without any** ASTERISKS ARE STATISTICALLY HIGHLY SIGNIFICANT WITH P<0.001. AND ** : P<0.01, AND * : P<0.05 , – : P>0.05 AND N.D. MEANS NO DATA.

	HsO	LCh	Lesk	WuP	Resnik	JCn	Lin	Path	vec_p	vec	C_A	C_P	C_V	C_{NPMI}	C_{UCI}	C_{UMass}
MC(P)	–	0.57*	–	0.55*	0.59	–	0.53*	–	0.60	**0.88**	–	0.79	–	0.77	0.67	–
MC(S)	–	0.58*	0.60	0.55*	0.68	–	0.56*	0.56*	0.70	**0.90**	–	0.81	0.65	0.82	–	–
RG(P)	0.54	0.60	0.44	0.53	0.61	–	0.54	0.54	n.d.	n.d.	–	0.75	–	**0.77**	0.71	–
RG(S)	0.49	0.56	0.55	0.51	0.55	–	0.46	0.54	n.d.	n.d.	–	**0.85**	0.50	0.84	0.83	0.45
Lau(P)	0.19	–	0.15	0.18	0.25	0.33	0.29	–	n.d.	n.d	0.38	**0.61**	0.31	0.55	0.51	0.28
Lau(S)	0.25	–	0.19	0.20	0.31	0.39	0.37	–	n.d.	n.d.	0.39	**0.52**	0.33	0.49	0.46	0.26
Simlex n.(P)	0.35	**0.52**	0.25	0.45	0.41	0.35	0.51	0.51	0.28	0.35	–	0.24	0.13	0.17	0.18	–
Simlex n.(S)	0.36	0.49	0.31	0.47	0.41	**0.51**	**0.51**	0.48	0.22	0.33	–	0.22	0.21	0.16	0.18	–

-.

% /- ,

;

$
%
= 7 8 8 8

' % B ?
C B
4--- + 0 8
!
D,--% ( )@