Word Association Measures for Finding Verb Sense Patterns

Patrick Hanks, Ismail El Maarouf, and Michael Oakes
patrick.w.hanks@gmail.com, i.el-maarouf@wlv.ac.uk, Michael.Oakes@wlv.ac.uk

Relation to PARSEME

- Use of statistical measures for associations between 3 variables to find multi-word units in the BNC50 Corpus.
- A hybrid approach which quantifies the collocational strength of dependency relations found by the Stanford Parser.

Pattern Dictionary of English Verb ('Act')

No.	%	Pattern / *implicatures*
1	16.4%	**[[{Human = Agent} | {Institution = Agent} | {Animal = Agent} | {Machine = Agent}]] act [NO OBJ]** *[[Human | Institution | Animal | Machine = Agent]] performs a motivated [[Action]]*
2	23.8%	**[[Human | Institution]] act [NO OBJ] [Adv[Manner]]** *[[Human | Institution]] behaves in the manner specified*
12	<0.01%	*phrasal verb* **[[Machine]] act [NO OBJ] {up}** *[[Machine]] fails to function correctly*
13	<0.01%	*idiom* **[[Human]] act {POSDET age}** *[[Human]] behaves in a manner appropriate to their age*

Church and Hanks proposed statistics for word associations

- Pointwise Mutual Information (PMI) between word x and word y is given by the formula

$$I(x,y) = log_2 \frac{P(x,y)}{P(x).P(y)}$$

Where P(x,y) is the probability of the two words occurring in a common context (such as a span of 5 words, or in subject-object relation),
while P(x) and P(y) are the probabilities of finding words x and y respectively anywhere in the corpus.

- PMI is positive if the two words tend to co-occur, 0 if they occur together as often as one would expect by chance, and less than 0 if they are in complementary distribution (Church and Hanks, 1989).
- PMI was used to examine the content word collocates of the verb 'shower', which were found to include 'abuse', 'accolades', 'affection', 'applause', and 'arrows'.

Association Measures for 3 variables (Van de Cruys, 2011)

Specific Correlation

$$SC(x,y,z) = log_2 \frac{P(x,y,z)}{P(x).P(y).P(z)}$$

Specific Interaction Information

$$SII(x,y,z) = log_2 \frac{P(x,y),P(y,z),P(x,z)}{P(x).P(y).P(z).P(x,y,z)}$$

SVO triples found by SC

Verb	Subject	Object	SC
SPARE	DEAR	RIB	34.81
WEEKEND	ENDGAMES	COMPETITION	34.39
OPENED	WHICH	JUNE	33.94
Pick	Hella	Editor	31.66
Added	Value	Tax	29.81
es	Dar	Salaam	29.55
See	4	Fig.	28.82
pine	p	marten	27.64
glazed	UPVC	window	25.82
spotted	Numbers	owl	25.70
assure	hon.	House	25.62
NEED	YOU	WHAT	25.54
branching	maximum	ratio	25.52
learned	hon.	Friend	24.32
addressed	stamped	envelope	24.29
firmed	shares	3p	23.95
charges	Girobank	fee	23.93

V	S	O	SC
bear	I	it	2.55
bear	he	it	1.71
bear	market	what	10.17
bear	that	name	7.56
bear	they	resemblance	9.15
bear	which	resemblance	8.92
bear	who	brunt	10.49
bears	it	resemblance	10.71
bears	that	name	7.99
bears	which	name	7.23
bears	which	resemblance	10.69

SVO triples found by SII

Verb	Subject	Object	Score
SPARE	DEAR	RIB	17.40
see	resignation	p.	17.04
WEEKEND	ENDGAMES	COMPETITION	16.99
see	elections	p.	16.75
OPENED	WHICH	JUNE	16.53
Compare	Us	Mythologies	15.58
es	Dar	Salaam	14.90
Pick	Hella	Editor	14.48
Added	Value	Tax	14.22
learned	hon.	Friend	12.76
assure	hon.	House	12.14
glazed	UPVC	window	11.00
pine	p	marten	10.82
NEED	YOU	WHAT	10.57
contact	Anyone	CID	9.72
starts	Sale	Tuesday	9.69
firmed	shares	3p	9.60

V	S	O	SC
carried	he	it	-1.73
carried	which	penalty	-3.66
carried	which	sentence	-4.72
carried	who	research	-2.59
carried	who	work	-1.60
carries	it	which	-1.43
carries	which	maximum	-1.58
carries	which	penalty	-1.36
carry	they	which	0.14

All triples found by SC

V	Arg1	Arg2	SC	Freq.
V=abandon	dobj=ourselves	prep=to=God	8.92	5
V=abandon	prep=to=principle	dobj=commitment	8.18	3
V=abandon	nsubj=scientist	dobj=theory	5.84	3
V=abandon	dobj=idea	advmod=altogether	4.56	3
V=abandon	conj=decide	dobj=idea	4.40	3
V=abandon	xcomp=build	dobj=plan	4.21	3
V=abandon	auxpass=be	nsubjpass=siege	4.07	4
V=abandon	aux=to	dobj=reprocessing	3.70	3
V=abandon	dobj=attempt	nsubj=government	3.31	4
V=abandon	prep=to=fate	aux=have	3.05	3
V=abandon	auxpass=be	nsubjpass=match	2.07	4
V=abandon	nsubjpass=doctrine	aux=have	2.05	3
V=abandon	cc=and	conj=opt	1.94	3
V=abandon	nsubjpass=doctrine	auxpass=be	1.45	4
V=abandon	aux=should	advmod=altogether	1.44	3
V=abandon	aux=should	dobj=attempt	1.04	3
V=abandon	aux=should	prep=in=favour	1.02	4
V=abandon	conj=adopt	cc=and	0.91	7
V=abandon	complm=that	prep=to=principle	0.86	3
V=abandon	nsubjpass=farm	auxpass=be	0.71	3

Idioms and phrasal verbs

Verb	Pat.	Arg.	Head	Particle	PV	IDIOM
abandon	5	O	ship		f	t
abdicate	2	O	role | responsibility for [[Anything]]		f	f
abhor	2	O	vacuum		f	t
abhor	2	S	nature		f	t
abide	3	A	Christ | God's love	in	f	f
abide	4	S	Christ | God | Lord		f	f
abound	2	A	number| amount	in | with	t	f
abstain	3	A	vote	from	f	f
abstain	3	A	vote	in	f	f
abstract	4	A		from	t	f
abuse	5	O	body		f	f
accede	2	A	throne	to	f	f
accept	3	O	responsibility for [[Eventuality | Entity]]		f	f
accept	6	O	resignation		f	f
accept	7	O	appointment | job		f	f
ache	2	S	heart | bone		f	f
acquit	2	A	well		f	f
act	10	A		out	t	f
act	11	A		up	t	f
act	12	A		up	t	f
act	13	O	age		f	t
add	6	A		together	t	f
add	6	A		up	t	f
add	7	A		to	t	f
add	7	A		up	t	f
add	8	A		to	t	f
add	8	A		up	t	f

Results for Idioms

Top n	Specific Correlation	Specific Interaction Information	Raw Frequency
1	1.9	1.4	2.4
5	6.4	5.4	4.2
10	7.8	6.6	6.8
20	12	7.3	8
All	18.2	18.2	18.2

Results for phrasal verbs

Top n	Specific Correlation	Specific Interaction Information	Raw Frequency
1	7.6	6.9	9.9
5	14.1	15	15.9
10	18	18.9	20.3
20	24.9	21.5	21.9
All	37.6	37.6	37.6

Acknowledgements

This research was partially supported by an AHRC grant [DVC, AH/J005940/1, 2012-2015]

References

- Tim Van de Cruys. 2011. Two Multivariate Generalizations of Pointwise Mutual Information. Disco 2011, 24 June 2011, Portland, OR.
- Dan Klein and Christopher D. Manning. 2003. Accurate Unlexicalized Parsing. Proceedings of the 41st Meeting of the Association for Computational Linguistics, pp. 423-430.
- Pattern Dictionary of English Verbs **http://pdev.org.uk**
- Kenneth W. Church and Patrick Hanks. 1989. Word Association Norms, Mutual Information and Lexicography. Proceedings of the 27th Meeting of the Association for Computational Linguistics: 78-83.