



Taking that Next Step: Methods for Automatic Extraction of Semantic Data from Large Corpora

Matthew Munson

University of Leipzig, Germany

munson@dh.uni-leipzig.de

Twitter: @sonofmun



What DH does well!



What DH does well!

- Produce Data!



What DH does well!

- Produce Data!
 - Perseus



What DH does well!

- Produce Data!
 - Perseus
 - TLG



What DH does well!

- Produce Data!
 - Perseus
 - TLG
 - Codex Sinaiticus Online



What DH does well!

- Produce Data!
 - Perseus
 - TLG
 - Codex Sinaiticus Online
 - Virtual Manuscript Room



What DH does well!

- Produce Data!
 - Perseus
 - TLG
 - Codex Sinaiticus Online
 - Virtual Manuscript Room
 - Driven by TEI



What DH doesn't do as well (yet)



What DH doesn't do as well (yet)

- Analyze this data to find meaning



What DH doesn't do as well (yet)

- Analyze this data to find meaning
- Humanities = finding meaning



What DH doesn't do as well (yet)

- Analyze this data to find meaning
- Humanities = finding meaning
- DH needs to take that next step!



What DH doesn't do as well (yet)

- Analyze this data to find meaning
- Humanities = finding meaning
- DH needs to take that next step!
 - Develop computational methods to help find meaning



What DH doesn't do as well (yet)

- Analyze this data to find meaning
- Humanities = finding meaning
- DH needs to take that next step!
 - Develop computational methods to help find meaning
 - Apply these methods to our data sources



What DH doesn't do as well (yet)

- Analyze this data to find meaning
- Humanities = finding meaning
- DH needs to take that next step!
 - Develop computational methods to help find meaning
 - Apply these methods to our data sources
 - Produce exciting humanities research results



This presentation



This presentation

- A modest proposal to take that step



This presentation

- A modest proposal to take that step
- Explores:



This presentation

- A modest proposal to take that step
- Explores:
 - Computational methods to extract semantic relationships



This presentation

- A modest proposal to take that step
- Explores:
 - Computational methods to extract semantic relationships
 - Their applications to ancient texts



This presentation

- A modest proposal to take that step
- Explores:
 - Computational methods to extract semantic relationships
 - Their applications to ancient texts
 - Change in word meaning between LXX and NT



This presentation

- A modest proposal to take that step
- Explores:
 - Computational methods to extract semantic relationships
 - Their applications to ancient texts
 - Change in word meaning between LXX and NT
 - Some initial results



Semantic Relationships



Semantic Relationships

- Distributional semantics



Semantic Relationships

- Distributional semantics
 - Firth, “You shall know a word by the company it keeps!”



Semantic Relationships

- Distributional semantics
 - Firth, “You shall know a word by the company it keeps!”
 - Harris, “If we consider words or morphemes A and B to be more different in meaning than A and C, then we will often find that the distributions of A and B are more different than the distributions of A and C. In other words, difference of meaning correlates with difference of distribution.”



Semantic Relationships

- Distributional semantics
 - Firth, “You shall know a word by the company it keeps!”
 - Harris, “If we consider words or morphemes A and B to be more different in meaning than A and C, then we will often find that the distributions of A and B are more different than the distributions of A and C. In other words, **difference of meaning correlates with difference of distribution.**”



An example



An example

- Word meaning differs where distribution differs



An example

- Word meaning differs where distribution differs
- What is the difference between “big” and “large”?



An example

- Word meaning differs where distribution differs
- What is the difference between “big” and “large”?
 - Big occurs frequently with “sister”, large does not



An example

- Word meaning differs where distribution differs
- What is the difference between “big” and “large”?
 - Big occurs frequently with “sister”, large does not
 - We see that “big” has a sense that “large” does not



Overview of Method



Overview of Method

- 1) Start with a lemmatized text of Greek OT and NT



Overview of Method

- 1) Start with a lemmatized text of Greek OT and NT
- 2) Count co-occurrences of word lemmata



Overview of Method

- 1) Start with a lemmatized text of Greek OT and NT
- 2) Count co-occurrences of word lemmata
 - Window of 4 left and 4 right of target word



Overview of Method

- 1) Start with a lemmatized text of Greek OT and NT
- 2) Count co-occurrences of word lemmata
 - Window of 4 left and 4 right of target word
- 3) Calculate statistical significance of co-occurrence



Overview of Method

- 1) Start with a lemmatized text of Greek OT and NT
- 2) Count co-occurrences of word lemmata
 - Window of 4 left and 4 right of target word
- 3) Calculate statistical significance of co-occurrence
- 4) Compare the words between and within the testaments



Note of Caution



Note of Caution

- Ultimate in historical-critical method



Note of Caution

- Ultimate in historical-critical method
 - meaning is relational



Note of Caution

- Ultimate in historical-critical method
 - meaning is relational
 - But this makes determining “actual meaning” a very complex process



Note of Caution

- Ultimate in historical-critical method
 - meaning is relational
 - But this makes determining “actual meaning” a very complex process
 - the “meaning” of every word is communicated by the meaning of every other word with which it co-occurs



Note of Caution

- Ultimate in historical-critical method
 - meaning is relational
 - But this makes determining “actual meaning” a very complex process
 - the “meaning” of every word is communicated by the meaning of every other word with which it co-occurs
 - and every word with which the co-occurents co-occur



Note of Caution

- Ultimate in historical-critical method
 - meaning is relational
 - But this makes determining “actual meaning” a very complex process
 - the “meaning” of every word is communicated by the meaning of every other word with which it co-occurs
 - and every word with which the co-occurents co-occur
 - etc.



Why Lemmatize?



Why Lemmatize?

- Biblical Greek is morphologically complex



Why Lemmatize?

- Biblical Greek is morphologically complex
 - Nouns – gender, number, case



Why Lemmatize?

- Biblical Greek is morphologically complex
 - Nouns – gender, number, case
 - Verbs – person, number, mood, voice, tense



Why Lemmatize?

- Biblical Greek is morphologically complex
 - Nouns – gender, number, case
 - Verbs – person, number, mood, voice, tense
- This atomizes data



Why Lemmatize?

- Biblical Greek is morphologically complex
 - Nouns – gender, number, case
 - Verbs – person, number, mood, voice, tense
- This atomizes data
 - e.g., information about God (θεός - 1307) in NT split among θεός (69), θεός (236), θεῶ (157), θεοῦ (687), θεόν (74), θεόν (73)



Why Lemmatize?

- Biblical Greek is morphologically complex
 - Nouns – gender, number, case
 - Verbs – person, number, mood, voice, tense
- This atomizes data
 - e.g., information about God (θεός - 1307) in NT split among θεός (69), θεός (236), θεῶ (157), θεοῦ (687), θεόν (74), θεόν (73)
- Biblical corpus is too small (~770,000 words)



Counting Co-occurrences: 4L-4R (1)



Counting Co-occurrences: 4L-4R (1)

- Why 4 left and 4 right?



Counting Co-occurrences: 4L-4R (1)

- Why 4 left and 4 right?
- Experimental answer varies:



Counting Co-occurrences: 4L-4R (1)

- Why 4 left and 4 right?
- Experimental answer varies:
 - Bullinaria - ~8L-8R for smaller corpora



Counting Co-occurrences: 4L-4R (1)

- Why 4 left and 4 right?
- Experimental answer varies:
 - Bullinaria - ~8L-8R for smaller corpora
 - León – 4L-4R



Counting Co-occurrences: 4L-4R (1)

- Why 4 left and 4 right?
- Experimental answer varies:
 - Bullinaria - ~8L-8R for smaller corpora
 - León – 4L-4R
 - Other studies – as low as 2L-2R



Counting Co-occurrences: 4L-4R (1)

- Why 4 left and 4 right?
- Experimental answer varies:
 - Bullinaria - ~8L-8R for smaller corpora
 - León – 4L-4R
 - Other studies – as low as 2L-2R
- So which to choose?



Counting Co-occurrences: 4L-4R (2)



Counting Co-occurrences: 4L-4R (2)

- From cognitive psychology



Counting Co-occurrences: 4L-4R (2)

- From cognitive psychology
- Miller's Magic Number: 7 ± 2



Counting Co-occurrences: 4L-4R (2)

- From cognitive psychology
- Miller's Magic Number: 7 ± 2
- More recent experiments: actually 3-4



Counting Co-occurrences: 4L-4R (2)

- From cognitive psychology
- Miller's Magic Number: 7 ± 2
- More recent experiments: actually 3-4
- What does this have to do with co-occurrences?



Counting Co-occurrences: 4L-4R (3)

the	small	brown	cat	jumped	the	lazy	black	dog



Counting Co-occurrences: 4L-4R (3)

the	small	brown	cat	jumped	the	lazy	black	dog



Counting Co-occurrences: 4L-4R (3)

4	3	2	1	***				
the	small	brown	cat	jumped	the	lazy	black	dog



Counting Co-occurrences: 4L-4R (3)

4	3	2	1	***				
the	small	brown	cat	jumped	the	lazy	black	dog

the	small	brown	cat	jumped	the	lazy	black	dog



Counting Co-occurrences: 4L-4R (3)

4	3	2	1	***				
the	small	brown	cat	jumped	the	lazy	black	dog

				4	3	2	1	***
the	small	brown	cat	jumped	the	lazy	black	dog



Counting Co-occurrences: 4L-4R (4)



Counting Co-occurrences: 4L-4R (4)

- Cognitively and experimentally, 4L-4R is a good starting place



Counting Co-occurrences: 4L-4R (4)

- Cognitively and experimentally, 4L-4R is a good starting place
- NB: it is only a starting place



Counting Co-occurrences: 4L-4R (4)

- Cognitively and experimentally, 4L-4R is a good starting place
- NB: it is only a starting place
- It needs to be tested by its results



Counting Co-occurrences: 4L-4R (4)

- Cognitively and experimentally, 4L-4R is a good starting place
- NB: it is only a starting place
- It needs to be tested by its results
- This is no different than the normal humanities research workflow



Counting Co-occurrences: 4L-4R (4)

- Cognitively and experimentally, 4L-4R is a good starting place
- NB: it is only a starting place
- It needs to be tested by its results
- This is no different than the normal humanities research workflow
- It is just made visible instead of hidden



Count Co-Occurrence

L4	L3	L2	L1	Target	R1	R2	R3	R4
έν	ἀρχή	ποιέω	ὁ	θεός	ὁ	οὐρανός	καί	ὁ
ὁ	ἄβυσσος	καί	πνεῦμα	θεός	ἐπιφέρω	ἐπάνω	ὁ	ὔδωρ



Count Co-Occurrence

L4	L3	L2	L1	Target	R1	R2	R3	R4
έν	ἀρχή	ποιέω	ὁ	θεός	ὁ	οὐρανός	καί	ὁ
ὁ	ἄβυσσος	καί	πνεῦμα	θεός	ἐπιφέρω	ἐπάνω	ὁ	ὔδωρ

Counts:

ὁ - 5



Count Co-Occurrence

L4	L3	L2	L1	Target	R1	R2	R3	R4
έν	ἀρχή	ποιέω	ὁ	θεός	ὁ	οὐρανός	καί	ὁ
ὁ	ἄβυσσος	καί	πνεῦμα	θεός	ἐπιφέρω	ἐπάνω	ὁ	ὔδωρ

Counts:

ὁ - 5

καί - 2



Count Co-Occurrence

L4	L3	L2	L1	Target	R1	R2	R3	R4
έν	ἀρχή	ποιέω	ὁ	θεός	ὁ	οὐρανός	καί	ὁ
ὁ	ἄβυσσος	καί	πνεῦμα	θεός	ἐπιφέρω	ἐπάνω	ὁ	ὔδωρ

Counts:

ὁ - 5

καί - 2

έν - 1

ἀρχή - 1

ποιέω - 1

ἄβυσσος - 1

πνεῦμα - 1

ἐπιφέρω - 1

ἐπάνω - 1

ὔδωρ - 1



Statistical Significance: Why



Statistical Significance: Why

- Notice the counts for the stop words (ὁ, καί)?



Statistical Significance: Why

- Notice the counts for the stop words (ὁ, καί)?
- If we used pure counts, ὁ would score the highest for (almost) every word



Statistical Significance: Why

- Notice the counts for the stop words (ὁ, καί)?
- If we used pure counts, ὁ would score the highest for (almost) every word
- This would make little sense, especially semantically



Statistical Significance: Why

- Notice the counts for the stop words (ὁ, καί)?
- If we used pure counts, ὁ would score the highest for (almost) every word
- This would make little sense, especially semantically
- So...



Calculate Statistical Significance



Calculate Statistical Significance

- Significant co-occurrence happens when two words occur together more often than you would expect by chance



Calculate Statistical Significance

- Significant co-occurrence happens when two words occur together more often than you would expect by chance
 - LXX: ὁ 1 every 7 words, κύριος 1 every 72



Calculate Statistical Significance

- Significant co-occurrence happens when two words occur together more often than you would expect by chance
 - LXX: ὁ 1 every 7 words, κύριος 1 every 72
 - w/ θεός: ὁ 1 every 5 words, κύριος 1 every 22



Calculate Statistical Significance

- Significant co-occurrence happens when two words occur together more often than you would expect by chance
 - LXX: ὁ 1 every 7 words, κύριος 1 every 72
 - w/ θεός: ὁ 1 every 5 words, κύριος 1 every 22
 - θεός attracts κύριος more than it attracts ὁ



Calculate Statistical Significance

- Significant co-occurrence happens when two words occur together more often than you would expect by chance
 - LXX: ὁ 1 every 7 words, κύριος 1 every 72
 - w/ θεός: ὁ 1 every 5 words, κύριος 1 every 22
 - θεός attracts κύριος more than it attracts ὁ
 - This would suggest that κύριος tells us more about the word θεός than ὁ does



Statistical Significance: How



Statistical Significance: How

- Log-likelihood ratio



Statistical Significance: How

- Log-likelihood ratio
- Returns a number that represents how much one word attracts another



Statistical Significance: How

- Log-likelihood ratio
- Returns a number that represents how much one word attracts another
- Positive means attraction, negative means repulsion



Statistical Significance: How

- Log-likelihood ratio
- Returns a number that represents how much one word attracts another
- Positive means attraction, negative means repulsion
- This should tell us how much one depends on the other to create meaning



Different Ways Forward



Different Ways Forward

- Compare LL lists manually, but...



Different Ways Forward

- Compare LL lists manually, but...
 - 3,893 words common to both testaments



Different Ways Forward

- Compare LL lists manually, but...
 - 3,893 words common to both testaments
 - Over 14,000 (LXX) or 5,300 (NT) possible co-occurrences



Different Ways Forward

- Compare LL lists manually, but...
 - 3,893 words common to both testaments
 - Over 14,000 (LXX) or 5,300 (NT) possible co-occurrences
 - Need to reduce the complexity



Different Ways Forward

- Compare LL lists manually, but...
 - 3,893 words common to both testaments
 - Over 14,000 (LXX) or 5,300 (NT) possible co-occurrences
 - Need to reduce the complexity
- Compare LL lists automatically with...



...Cosine Similarity



...Cosine Similarity

- I calculated similarity between the testaments



...Cosine Similarity

- I calculated similarity between the testaments
 - e.g., how similar are the log-likelihood lists for θεός in the OT and θεός in the NT



...Cosine Similarity

- I calculated similarity between the testaments
 - e.g., how similar are the log-likelihood lists for θεός in the OT and θεός in the NT
 - This tells us which words have changed the most between the testaments



...Cosine Similarity

- I calculated similarity between the testaments
 - e.g., how similar are the log-likelihood lists for θεός in the OT and θεός in the NT
 - This tells us which words have changed the most between the testaments
- And within the testaments



...Cosine Similarity

- I calculated similarity between the testaments
 - e.g., how similar are the log-likelihood lists for θεός in the OT and θεός in the NT
 - This tells us which words have changed the most between the testaments
- And within the testaments
 - e.g., how similar is the log-likelihood list for θεός in the OT with that of κύριος in the OT



...Cosine Similarity

- I calculated similarity between the testaments
 - e.g., how similar are the log-likelihood lists for θεός in the OT and θεός in the NT
 - This tells us which words have changed the most between the testaments
- And within the testaments
 - e.g., how similar is the log-likelihood list for θεός in the OT with that of κύριος in the OT
 - Tells us which words are most similar within each testament



And then...



And then...

- Investigate change in usage (log-likelihood lists)



And then...

- Investigate change in usage (log-likelihood lists)
 - OT Log Likelihood – NT Log Likelihood = change



And then...

- Investigate change in usage (log-likelihood lists)
 - OT Log Likelihood – NT Log Likelihood = change
- Investigate change in meaning (cosine similarity lists)



And then...

- Investigate change in usage (log-likelihood lists)
 - OT Log Likelihood – NT Log Likelihood = change
- Investigate change in meaning (cosine similarity lists)
 - OT COS Similarity – NT COS Similarity = change



And then...

- Investigate change in usage (log-likelihood lists)
 - $OT \text{ Log Likelihood} - NT \text{ Log Likelihood} = \text{change}$
- Investigate change in meaning (cosine similarity lists)
 - $OT \text{ COS Similarity} - NT \text{ COS Similarity} = \text{change}$
- Here's how...



1. Determine Difference in Usage

Word	Similarity	Occurrences OT	Occurrences NT
Σίμων	0.124978237	80	75
χάρις	0.125823442	134	155
γεννάω	0.177162603	254	97
ἐκβάλλω	0.178899156	101	81
προσεύχομαι	0.258272324	110	85
ἕτερος	0.268276562	267	98
Μωϋσῆς	0.279342059	826	80
ἄλλος	0.280188682	108	155
διδάσκω	0.282892057	107	97
ἑπτὰ	0.292734261	377	88
καλός	0.296850637	235	102
νεκρός	0.298957827	90	128
ἔτι	0.316743475	550	93
οὐδείς	0.317224244	188	234
ἐγείρω	0.322611383	80	144
ἐκκλησία	0.334642299	103	114
βλέπω	0.34332697	133	132
ζωή	0.349299071	298	135
προσέρχομαι	0.374945203	116	86
πονηρός	0.38284153	382	78



1. Determine Difference in Usage

Word	Similarity	Occurrences OT	Occurrences NT
Σίμων	0.124978237	80	75
χάρις	0.125823442	134	155
γεννάω	0.177162603	254	97
ἐκβάλλω	0.178899156	101	81
προσεύχομαι	0.258272324	110	85
ἕτερος	0.268276562	267	98
Μωϋσῆς	0.279342059	826	80
ἄλλος	0.280188682	108	155
διδάσκω	0.282892057	107	97
ἑπτά	0.292734261	377	88
καλός	0.296850637	235	102
νεκρός	0.298957827	90	128
ἔτι	0.316743475	550	93
οὐδεὶς	0.317224244	188	234
ἐγείρω	0.322611383	80	144
ἐκκλησία	0.334642299	103	114
βλέπω	0.34332697	133	132
ζωή	0.349299071	298	135
προσέρχομαι	0.374945203	116	86
πονηρός	0.38284153	382	78



2. Compare Change in Usage

Σίμων - Simon

Collocate	Log-likelihood _{Diff.}	English Translation
ἀρχιερέυς	4.78868755	High Priest
ἀδελφός	1.97668932	Brother
σύ	0.70884314	You
δέ	0.61095764	But, and
μέγας	0.46820919	Great
ἐγώ	0.44210074	I
καλέω	0.29046041	To call
τις	0.20737781	Someone
αἴρω	0.20060086	To lift up
...		
θάλασσα	-0.35265681	Sea
ἀποκρίνομαι	-0.37932464	To set apart
πρός	-0.39142971	Towards
οὖν	-0.43430816	And so
λέγω	-0.80210635	To say
ὁ	-0.940275	The
ἐπικαλέω	-0.98102451	Summon
καί	-1.06836515	And
Ἰούδας	-1.96547425	Judas

χάρις - grace, favor

Collocate	Log-likelihood _{Diff.}	English Translation
εὕρισκω	17.02945	To find
ἐναντίον	7.104399	Opposite
ὁ	4.55197	The
ἐνώπιον	2.082461	Facing
εἰ	1.887208	If
ἀγαθός	1.432463	Good
δόξα	0.988545	Glory
παρά	0.711348	Beside
σύ	0.580478	You
...		
σοφία	-0.53002	Wisdom
ἔλεος	-0.68112	Mercy
διά	-0.91597	Through
κατά	-0.9759	Downwards
κύριος	-0.99528	Lord
καί	-1.14771	And
αὐτός	-1.53347	Himself
λέγω	-1.59654	To say
θεός	-3.62867	God



2. Change in Usage Analysis

Σίμων

- OT
 - appears with ἀρχιερεύς and ἀδελφός (he was high priest and brother of Judas and Jonathan Maccabee)
- NT
 - appears with Ἰούδας (“Judas son of Simon”), with ἐπικαλέω and ἀποκρίνομαι (apostolic words) and with θάλασσα (the sea)

χάρις

- OT
 - appears with εὕρισκω (to find), with ἐναντίον and ἐνώπιον (facing words), and with ἀγαθός (description of gifts)
- NT
 - appears with θεός and κύριος (giver of gifts) and σοφία and ἔλεος (possible gifts)



3. Compare Change in Topic

Σίμων

Similar Word	Cosine Similarity Difference	English
πολεμέω	0.354931673	to fight, to make war
Σαλωμών	0.354820728	Solomon
έξολεθρεύω	0.347724511	to destroy utterly
νέος	0.34625518	youthful, new
καταλείπω	0.345432626	to leave behind
Ονίας	0.337392476	Onias
ίςχύς	0.336073321	strength, might
Ιωναθαν	0.334101138	Jonathan
πατάσσω	0.334069924	strike, smite
...		
Πέτρος	-0.306455693	Peter
άνακαθίζω	-0.331895269	to sit up
διενθυμέομαι	-0.33809918	ponder
νεύω	-0.343399149	to nod
Ίάκωβος	-0.369768608	James
ξενίζω	-0.402026864	to entertain, astonish
φράζω	-0.418118756	to explain, interpret
άλιεύω	-0.449737419	to fish
Άνδρέας	-0.461752972	Andrew

χάρις

Similar Word	Cosine Similarity Difference	English
κατασχίζω	0.559569675	to cut up, to tear up
παροινέω	0.535813921	to insult, to rage against
συνοδία	0.531470401	caravan, company of travellers
κόνδυ	0.485367609	drinking vessel, cup
νεφθαί	0.437069069	naphtha
άντρώδης	0.390486532	cave-like
καρπώω	0.385592739	to offer (as a burnt offering)
διόρυγμα	0.379334876	digging through, breach, hole
δεκαπλασίων	0.374839561	ten-fold
...		
οὐν	-0.553553397	and so
ός	-0.553573814	who, which, what, that
γάρ	-0.56084709	for
θεός	-0.561622448	God
λέγω	-0.562754892	to say
δέ	-0.574144999	and, but
ίνα	-0.575720796	so that
διά	-0.584088488	through
μαθητής	-0.668627965	pupil, apprentice, disciple



3. Change in Topic Analysis

Σίμων

- OT
 - shares same topic with πολεμέω, ἐξολεθρεύω, ἰσχύς, and πατάσσω (war words), and Ονίας and Ιωναθαν (two other high priests)
- NT
 - shares same topic with Ἄνδρέας, Ἰάκωβος, and Πέτρος (names of apostles), with ἀλιεύω (to fish), and with διενθυμέομαι and φράζω (interpretation words)

χάρις

- OT
 - shares same topic with κατασχίζω, κόνδου, νεφθαι, καρπόω (sacrificial words)
- NT
 - shares same topic with μαθητής, θεός, and λέγω. Words 10 and 11 are ἀγάπη and πίστις (gifts)



1. Determine Difference in Usage

Word	CSim Score OT to NT	Occ OT	Occ NT
ἀκούω	0,845822895	1067	428
ἄνθρωπος	0,849351607	1427	550
κύριος	0,857196676	8605	717
ἐκ	0,859703706	3831	914
ὄραω	0,863501359	1503	483
μετά	0,863538079	2522	469
θεός	0,863710763	4008	1317
γινώσκω	0,865117244	754	222
εἰσέρχομαι	0,86794173	710	194
ἡμέρα	0,869471581	2572	389
ἔρχομαι	0,874266188	1082	632
ποιέω	0,877562927	3387	568
δέ	0,880434567	4907	2792



1. Determine Difference in Usage

Word	Csim Score OT to NT	Occ OT	Occ NT
ἀκούω	0,845822895	1067	428
ἄνθρωπος	0,849351607	1427	550
κύριος	0,857196676	8605	717
ἐκ	0,859703706	3831	914
ὄραω	0,863501359	1503	483
μετά	0,863538079	2522	469
θεός	0,863710763	4008	1317
γινώσκω	0,865117244	754	222
εἰσέρχομαι	0,86794173	710	194
ἡμέρα	0,869471581	2572	389
ἔρχομαι	0,874266188	1082	632
ποιέω	0,877562927	3387	568
δέ	0,880434567	4907	2792



2. Compare Change in Usage

Septuagint Co-Occurrents

Collocate	LL Difference	Translation
κύριος	75,91009775	Lord
ἐγώ	21,09482382	I
σύ	20,60507202	You
λατρεύω	9,549640054	To Serve
εὐλογητός	9,037180715	Praise
ὅτι	8,880056787	That, Which
υἱός	8,106906643	Son
Ἰσραήλ	6,608892501	Israel
εὐλογέω	6,016550457	To Praise
ἕτερος	5,368952049	Other
άνήρ	4,369749152	Man
φοβέω	3,552670146	To Fear
σωτήρ	3,426153263	Saviour
βασιλεύς	3,386998461	King
ἐγκαταλείπω	3,333623505	To Forsake

New Testament Co-Occurrents

Collocate	LL Difference	Translation
ἐκεῖνος	-0,93287157	That One
έν	-1,278667122	In
ἔρχομαι	-1,308490266	To Come
διά	-1,315025651	Through
Χριστός	-1,681830995	Christ
εἰρήνη	-1,702872647	Peace
ένώπιον	-2,070345268	Before, Facing
θέλημα	-2,358967237	Will
εὐχαριστέω	-2,608624736	To Be Thankful
δοξάζω	-3,501485566	To Glorify
χάρις	-3,713703025	Grace, Gift
βασιλεία	-7,380142984	Kingdom
αὐτός	-12,12970823	Himself
καί	-32,60915365	And
ὁ	-116,551495	The



2. Compare Change in Usage

Septuagint Verbs

Collocate	LL Difference	Translation
κύριος	75,91009775	Lord
ἐγώ	21,09482382	I
σύ	20,60507202	You
λατρεύω	9,549640054	To Serve
εὐλογητός	9,037180715	Praise
ὅτι	8,880056787	That, Which
υἱός	8,106906643	Son
Ἰσραήλ	6,608892501	Israel
εὐλογέω	6,016550457	To Praise
ἕτερος	5,368952049	Other
άνήρ	4,369749152	Man
φοβέω	3,552670146	To Fear
σωτήρ	3,426153263	Saviour
βασιλεύς	3,386998461	King
ἐγκαταλείπω	3,333623505	To Forsake

New Testament Verbs

Collocate	LL Difference	Translation
ἐκεῖνος	-0,93287157	That One
ἐν	-1,278667122	In
ἔρχομαι	-1,308490266	To Come
διά	-1,315025651	Through
Χριστός	-1,681830995	Christ
εἰρήνη	-1,702872647	Peace
ἐνώπιον	-2,070345268	Before, Facing
θέλημα	-2,358967237	Will
εὐχαριστέω	-2,608624736	To Be Thankful
δοξάζω	-3,501485566	To Glorify
χάρις	-3,713703025	Grace, Gift
βασιλεία	-7,380142984	Kingdom
αὐτός	-12,12970823	Himself
καί	-32,60915365	And
ὁ	-116,551495	The



2. Compare Change in Usage

Septuagint Nouns

Collocate	LL Difference	Translation
κύριος	75,91009775	Lord
ἐγώ	21,09482382	I
σύ	20,60507202	You
λατρεύω	9,549640054	To Serve
εὐλογητός	9,037180715	Praise
ὅτι	8,880056787	That, Which
υἱός	8,106906643	Son
Ἰσραήλ	6,608892501	Israel
εὐλογέω	6,016550457	To Praise
ἕτερος	5,368952049	Other
άνήρ	4,369749152	Man
φοβέω	3,552670146	To Fear
σωτήρ	3,426153263	Saviour
βασιλεύς	3,386998461	King
ἐγκαταλείπω	3,333623505	To Forsake

New Testament Nouns

Collocate	LL Difference	Translation
ἐκεῖνος	-0,93287157	That One
ἐν	-1,278667122	In
ἔρχομαι	-1,308490266	To Come
διά	-1,315025651	Through
Χριστός	-1,681830995	Christ
εἰρήνη	-1,702872647	Peace
ἐνώπιον	-2,070345268	Before, Facing
θέλημα	-2,358967237	Will
εὐχαριστέω	-2,608624736	To Be Thankful
δοξάζω	-3,501485566	To Glorify
χάρις	-3,713703025	Grace, Gift
βασιλεία	-7,380142984	Kingdom
αὐτός	-12,12970823	Himself
καί	-32,60915365	And
ὁ	-116,551495	The



3. Compare Change in Topic

Septuagint

Similar Word	OT Sim	NT Sim	Sim Diff	Translation
Σαλωμών	0,690	0,000	0,690	Solomon
ἐξολεθρεύω	0,736	0,050	0,686	destroy utterly
πρόσταγμα	0,682	0,000	0,682	command
παιδάριον	0,681	0,004	0,677	child
πεδῖον	0,674	0,000	0,674	field
αἰχμαλωσία	0,629	-0,033	0,661	captivity
ἐγκαταλείπω	0,713	0,077	0,635	To forsake
ἀλλόφυλος	0,666	0,036	0,630	foreign
κτῆνος	0,615	-0,014	0,629	cattle
ἐναντίον	0,766	0,137	0,629	facing
ἐκζητέω	0,773	0,145	0,628	seek out
χριστός	0,612	0,000	0,612	anointed
ἐλπίζω	0,318	-0,294	0,611	To hope
καταλαμβάνω	0,588	-0,023	0,611	To lay hold of
ἐξάιρω	0,649	0,039	0,610	to drive away
Βενιαμίν	0,546	-0,050	0,597	Benjamin
ἐναντι	0,638	0,050	0,588	before
συντελέω	0,740	0,155	0,585	to finish
κακία	0,687	0,107	0,580	wickedness
πατάσσω	0,734	0,155	0,579	to strike

New Testament

Similar Word	OT Sim	NT Sim	Sim Diff	Translation
Σατανᾶς	0,000	0,419	-0,419	Satan
Φαρισαῖος	0,000	0,429	-0,429	Pharisees
μέλλω	0,256	0,698	-0,442	to be about to
πιστεύω	0,441	0,883	-0,442	to believe
ἐκατοντάρχης	-0,188	0,256	-0,444	centurion
ὕπαγω	0,064	0,511	-0,447	to go, die
τοιούτος	0,096	0,545	-0,449	like this
πίστις	0,373	0,829	-0,456	Belief
εὐαγγέλιον	0,025	0,490	-0,465	gospel
Πέτρος	0,000	0,494	-0,494	Peter
Πιλάτος	0,000	0,495	-0,495	Pilate
Χριστός	0,000	0,520	-0,520	Christ
οὖν	0,302	0,833	-0,532	and so
χάρις	0,138	0,699	-0,562	gift
ἵνα	0,339	0,904	-0,565	in order to
φανερῶ	-0,018	0,558	-0,577	To show
ἀπόστολος	0,000	0,613	-0,613	apostle
ἀγάπη	-0,008	0,639	-0,647	love
Παῦλος	0,000	0,662	-0,662	Paul
μαθητής	0,000	0,783	-0,783	Disciple



3. Change in Topic – LXX Similar Words

Chi-Squared, no limit			Chi-Squared Less Than .25			Chi-squared Less than .10		
Σαλωμών	0,690	Solomon	ἐκζητέω	0,916	seek out	καταλείπω	0,883	to forsake
ἐξολεθρεύω	0,686	destroy utterly	δείκνυμι	0,890	to show	χώρα	0,879	field
πρόσταγμα	0,682	command	πρότερος	0,886	earlier	ἁγιάζω	0,879	to sanctify
παιδάριον	0,677	child	τάσσω	0,882	to arrange	ἐτοιμάζω	0,879	to prepare
πεδῖον	0,674	field	ὀχύρωμα	0,877	stronghold	συντελέω	0,877	to finish
αἰχμαλωσία	0,661	captivity	τύπτω	0,873	to strike	Αἰγύπτιος	0,877	Egypt
ἐγκαταλείπω	0,635	to desert	πάρειμι	0,873	to be present	δείκνυμι	0,875	to show
ἄλλόφυλος	0,630	foreign	ἐφοράω	0,873	to watch over	πλανάω	0,875	to wander
κτῆνος	0,629	cattle	μνημόσυνον	0,873	memorial, memory	Αβιμελεχ	0,875	Abimelech
ἐναντίον	0,629	facing	ἔσθω	0,873	eat	ἐκζητέω	0,875	seek out
ἐκζητέω	0,628	to search for	ἐπιτάσσω	0,873	to command	ἅμα	0,875	together
χριστός	0,612	anointed	ἐξουθενώω	0,873	treat with contempt	πλατύς	0,875	wide
ἐλπίζω	0,611	hope	διάνοια	0,873	mind, heart	ἀναιρέω	0,875	to destroy
καταλαμβάνω	0,611	to take	μεθύσκω	0,873	to make drunk	ἔνδοξος	0,875	Honoured
ἐξαίρω	0,610	to drive away	κατέχω	0,873	to hold, detain	ῥάβδος	0,875	scepter
Βενιαμίν	0,597	Benjamin	εὐδοκέω	0,873	to take delight	μεγαλύνω	0,875	To magnify
ἔναντι	0,588	opposite, before	ἀριθμέω	0,873	to count	κατάσχεσις	0,875	Possession
συντελέω	0,585	to finish	ἐκτρίβω	0,873	to destroy	θεωρέω	0,875	to observe
κακία	0,580	wickedness	Ἰδουμαία	0,873	Idumea	κριτής	0,875	a judge, ruler
πατάσσω	0,579	to strike	συναντάω	0,873	to meet, happen	ἀπώλεια	0,875	destruction



3. Change in Topic – LXX Similar Words

Chi-Squared, no limit		
Σαλωμών	0,690	Solomon
ἐξολεθρεύω	0,686	destroy utterly
πρόσταγμα	0,682	command
παιδάριον	0,677	child
πεδῖον	0,674	field
αἰχμαλωσία	0,661	captivity
ἐγκαταλείπω	0,635	to desert
ἀλλόφυλος	0,630	foreign
κτῆνος	0,629	cattle
ἐναντίον	0,629	facing
ἐκζητέω	0,628	to search for
χριστός	0,612	anointed
ἐλπίζω	0,611	hope
καταλαμβάνω	0,611	to take
ἐξαίρω	0,610	to drive away
Βενιαμίν	0,597	Benjamin
ἐναντι	0,588	opposite, before
συντελέω	0,585	to finish
κακία	0,580	wickedness
πατάσσω	0,579	to strike

Chi-Squared Less Than .25		
ἐκζητέω	0,916	seek out
δείκνυμι	0,890	to show
πρότερος	0,886	earlier
τάσσω	0,882	to arrange
ὄχυρωμα	0,877	stronghold
τύπτω	0,873	to strike
πάρειμι	0,873	to be present
ἐφοράω	0,873	to watch over
μνημόσυνον	0,873	memorial, memory
ἔσθω	0,873	eat
ἐπιτάσσω	0,873	to command
ἐξουθενώω	0,873	treat with contempt
διάνοια	0,873	mind, heart
μεθύσκω	0,873	to make drunk
κατέχω	0,873	to hold, detain
εὐδοκέω	0,873	to take delight
ἀριθμέω	0,873	to count
ἐκτρίβω	0,873	to destroy
Ἰδουμαία	0,873	Idumea
συναντάω	0,873	to meet, happen

Chi-squared Less than .10		
καταλείπω	0,883	to forsake
χώρα	0,879	field
ἁγιάζω	0,879	to sanctify
ἐτοιμάζω	0,879	to prepare
συντελέω	0,877	to finish
Αἰγύπτιος	0,877	Egypt
δείκνυμι	0,875	to show
πλανάω	0,875	to wander
Αβιμελεχ	0,875	Abimelech
ἐκζητέω	0,875	seek out
ἅμα	0,875	together
πλατύς	0,875	wide
ἀναιρέω	0,875	to destroy
ἔνδοξος	0,875	honoured
ῥάβδος	0,875	scepter
μεγαλύνω	0,875	To magnify
κατάσχεσις	0,875	Possession
θεωρέω	0,875	to observe
κριτής	0,875	a judge, ruler
ἀπώλεια	0,875	destruction



3. Change in Topic – NT Similar Words

Chi-Squared, no limit		
Σατανᾶς	-0,419	Satan
Φαρισαῖος	-0,429	Pharisees
μέλλω	-0,442	to be about to
πιστεύω	-0,442	to believe
ἑκατοντάρχης	-0,444	centurion
ὑπάγω	-0,447	to go (away)
τοιούτος	-0,449	like this
πίστις	-0,456	trust, faith
εὐαγγέλιον	-0,465	gospel
Πέτρος	-0,494	Peter
Πιλάτος	-0,495	Pilate
Χριστός	-0,520	Christ
οὖν	-0,532	and so
χάρις	-0,562	grace, gift
ἵνα	-0,565	in order to
φανερῶ	-0,577	reveal, show
ἀπόστολος	-0,613	apostle
ἀγάπη	-0,647	love
Παῦλος	-0,662	Paul
μαθητής	-0,783	disciple

Chi-squared less than .25		
θαυμάζω	-0,671	to wonder
διάβολος	-0,685	devil
λοιπός	-0,691	remaining
πτωχός	-0,692	poor, needy
τέλος	-0,700	the end
μνημεῖον	-0,732	monument
ἀγάπη	-0,756	love
σταυρῶ	-0,762	to crucify
ἀπόστολος	-0,763	apostle
Παῦλος	-0,777	Paul
τοιούτος	-0,827	like this, certain
μαθητής	-0,832	disciple
μάρτυς	-0,967	witness
διό	-0,967	therefore
φανερῶ	-0,967	reveal, show
Σατανᾶς	-0,967	devil
βλασφημέω	-0,967	to blaspheme
εὐαγγελίζω	-0,967	to evangelize
ὀφείλω	-0,967	to owe
ὁμοίως	-0,967	likewise

Chi-squared less than .10		
ἄλλοτε	-0,563	other
μαρτυρέω	-0,570	to witness
οὖν	-0,587	and so, therefore
ἐμός	-0,590	mine, of me
ἵνα	-0,621	in order to
εὐαγγέλιον	-0,623	good news, gospel
βάλλω	-0,633	to drive out
ώρα	-0,646	time of day, hour
πέμπω	-0,670	send
ἄλλοτε	-0,685	more
χάρις	-0,764	grace, gift
ὄχλος	-0,840	crowd
μαθητής	-0,841	disciple
Παῦλος	-0,855	Paul
πίστις	-0,973	belief, trust
ὅπου	-0,973	where
μέλλω	-0,973	to be about to
ἀγάπη	-0,973	love
ἀπόστολος	-0,973	apostle
περιπατέω	-0,973	to behave, live



3. Change in Topic – NT Similar Words

Chi-Squared, no limit		
Σατανᾶς	-0,419	Satan
Φαρισαῖος	-0,429	Pharisees
μέλλω	-0,442	to be about to
πιστεύω	-0,442	to believe
ἑκατοντάρχης	-0,444	centurion
ὑπάγω	-0,447	to go (away)
τοιούτος	-0,449	like this
πίστις	-0,456	trust, faith
εὐαγγέλιον	-0,465	gospel
Πέτρος	-0,494	Peter
Πιλάτος	-0,495	Pilate
Χριστός	-0,520	Christ
οὖν	-0,532	and so
χάρις	-0,562	grace, gift
ἵνα	-0,565	in order to
φανερῶ	-0,577	reveal, show
ἀπόστολος	-0,613	apostle
ἀγάπη	-0,647	love
Παῦλος	-0,662	Paul
μαθητής	-0,783	disciple

Chi-squared less than .25		
θαυμάζω	-0,671	to wonder
διάβολος	-0,685	devil
λοιπός	-0,691	remaining
πτωχός	-0,692	poor, needy
τέλος	-0,700	the end
μνημεῖον	-0,732	monument
ἀγάπη	-0,756	love
σταυρῶ	-0,762	to crucify
ἀπόστολος	-0,763	apostle
Παῦλος	-0,777	Paul
τοιούτος	-0,827	like this, certain
μαθητής	-0,832	disciple
μάρτυς	-0,967	witness
διό	-0,967	therefore
φανερῶ	-0,967	reveal, show
Σατανᾶς	-0,967	devil
βλασφημέω	-0,967	to blaspheme
εὐαγγελίζω	-0,967	to evangelize
ὀφείλω	-0,967	to owe
ὁμοίως	-0,967	likewise

Chi-squared less than .10		
ἕτερος	-0,563	other
μαρτυρέω	-0,570	to witness
οὖν	-0,587	and so, therefore
ἐμός	-0,590	mine, of me
ἵνα	-0,621	in order to
εὐαγγέλιον	-0,623	good news, gospel
βάλλω	-0,633	to drive out
ώρα	-0,646	time of day, hour
πέμπω	-0,670	send
μᾶλλον	-0,685	more
χάρις	-0,764	grace, gift
ὄχλος	-0,840	crowd
μαθητής	-0,841	disciple
Παῦλος	-0,855	Paul
πίστις	-0,973	belief, trust
ὅπου	-0,973	where
μέλλω	-0,973	to be about to
ἀγάπη	-0,973	love
ἀπόστολος	-0,973	apostle
περιπατέω	-0,973	to behave, live



Conclusions about the Bible



Conclusions about the Bible

- Method produces a lot of noise



Conclusions about the Bible

- Method produces a lot of noise
- But it produces classifiable, logical results



Conclusions about the Bible

- Method produces a lot of noise
- But it produces classifiable, logical results
 - LXX – God is represented in many ways (ruling, war, seeking, Exodus)



Conclusions about the Bible

- Method produces a lot of noise
- But it produces classifiable, logical results
 - LXX – God is represented in many ways (ruling, war, seeking, Exodus)
 - NT – God's servants and their relationship to God



Conclusions about the Bible

- Method produces a lot of noise
- But it produces classifiable, logical results
 - LXX – God is represented in many ways (ruling, war, seeking, Exodus)
 - NT – God's servants and their relationship to God
 - NT seems much more uniform than the LXX (Jon Levenson)



Conclusions about the Method



Conclusions about the Method

- We have not shown any ground-breaking results



Conclusions about the Method

- We have not shown any ground-breaking results
- But in the cases we have looked at, it has shown us some things we might expect



Conclusions about the Method

- We have not shown any ground-breaking results
- But in the cases we have looked at, it has shown us some things we might expect
- This might give us the confidence to trust results that are not what we expect



Conclusions about the Method

- We have not shown any ground-breaking results
- But in the cases we have looked at, it has shown us some things we might expect
- This might give us the confidence to trust results that are not what we expect
- The road in front of us is long, but it is certainly time to take that first step on this road

Thank You!