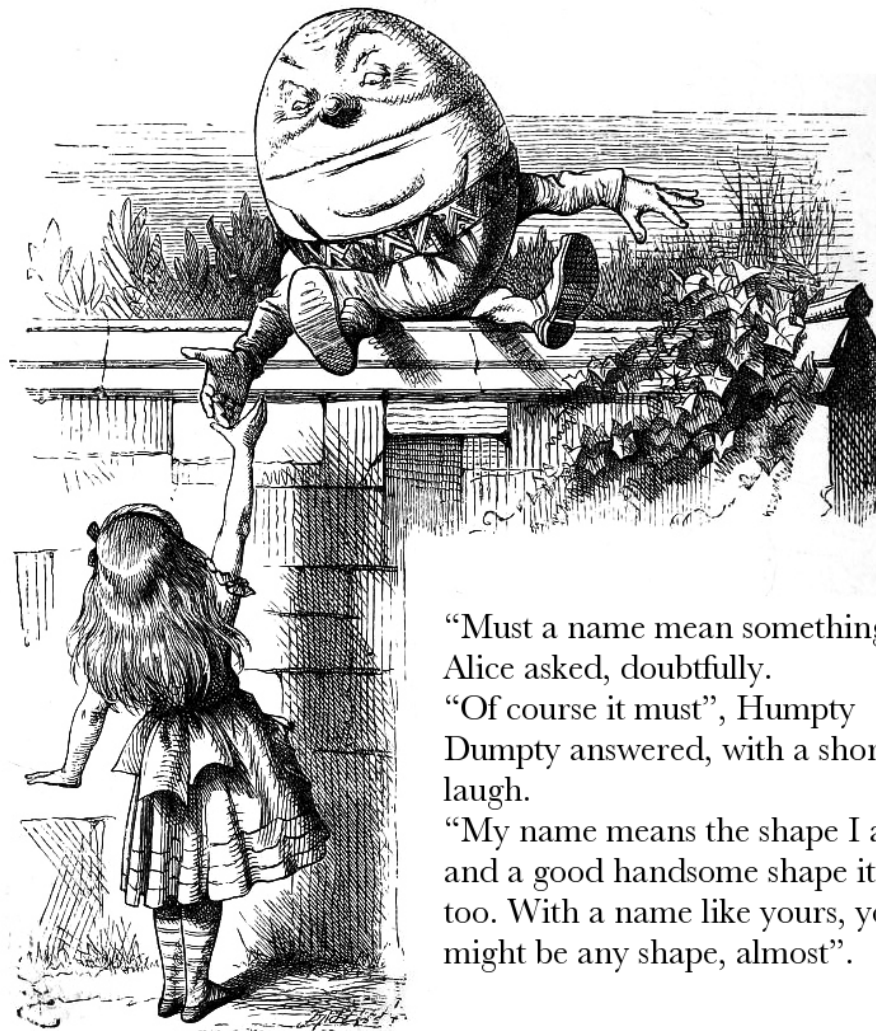


# What's in a name?

Jon Riding &  
Neil Boulton

United Bible Societies



“Must a name mean something”?  
Alice asked, doubtfully.

“Of course it must”, Humpty  
Dumpty answered, with a short  
laugh.

“My name means the shape I am,  
and a good handsome shape it is,  
too. With a name like yours, you  
might be any shape, almost”.



# Setting the scene

- Characteristics of Proper-Names (PNs)
  - Not translated (usually)
  - Particular to cultures
  - Often carry culture dependent meanings
  - Very difficult for a computer to identify

*Proper names are like poetry in the raw. Like all poetry they are untranslatable.*

W.H. Auden, "Names, Proper."  
A Certain World (1970).



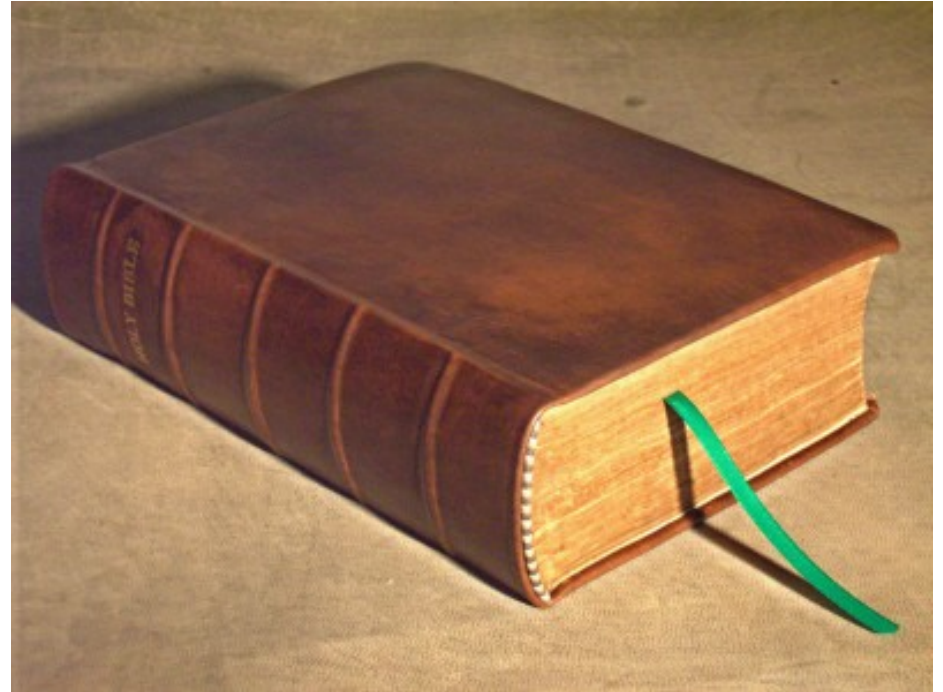
# Bible Translation in Context 1

- c. 7,000 active languages
- c. 550 completed translations of the Bible
- c. 3,500 languages with 'some portion'
- c. 3,500 without a translation
  - c. 1.5b people



# Bible Translation in Context 2

- A collection of between 66 and 81 books
- Many different genres
  - History, Prophecy, Poetry, Wisdom, Mythology, Biography, Correspondence, Apocalypse
- Translations typically take 10-15 years
- Often 'staged':
  - New Testament & Psalms,
  - Old Testament,
  - Deutero-Canonicals.



# Bible Translation in Context 3

- Over a 10-15 year project there can be many changes and not just with the text
  - Personnel and technology may change
  - Linguistic use may develop
  - Competencies improve
- These and more all lead to challenges in consistency



# Managing Key Terms

- Categories
  - Flora, Fauna, Measurements, Ritual, etc... and
  - 5,000 Proper Names
    - People, tribes, nations, clans, places, regions etc...

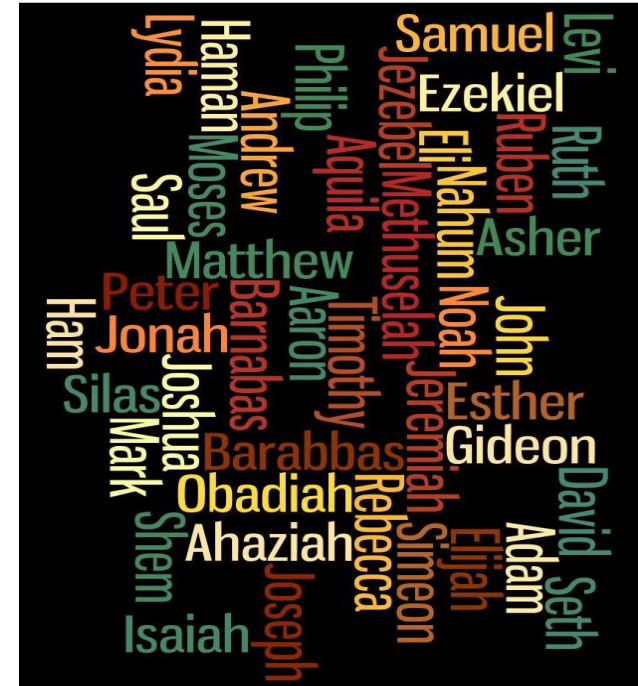
```
<Term Id="אַבִּי">
  <Transliteration>'ăbî</Transliteration>
  <Category>PN</Category>
  <Domain>person</Domain>
  <Language>hebrew</Language>
  <Definition>daughter of Zechariah; wife of Ahaz
  <Gloss>Abi</Gloss>
  <References>
    <Verse>2KI 18:2</Verse>
  </References>
</Term>
<Term Id="אַבֵּי הָעֶזְרִי">
  <Transliteration>'ăbiy hă'ezrî</Transliteration>
  <Category>PN</Category>
  <Domain>group</Domain>
  <Language>hebrew</Language>
  <Definition>member of clan of Abiezer</Definition>
  <Gloss>Abiezrite</Gloss>
  <References>
    <Verse>JDG 6:11</Verse>
    <Verse>JDG 6:24</Verse>
    <Verse>JDG 8:32</Verse>
  </References>
</Term>
<Term Id="אַבִּי-עֲלֹבֹן">
  <Transliteration>'ăbî-'ălbôn</Transliteration>
  <Category>PN</Category>
  <Domain>person</Domain>
  <Language>hebrew</Language>
  <Definition>warrior in time of king David; also
  <Gloss>Abi-Albon</Gloss>
  <References>
    <Verse>2SA 23:31</Verse>
  </References>
</Term>
<Term Id="אַבִּי-אֵל-1">
  <Transliteration>'ăbî'ēl-1</Transliteration>
  <Category>PN</Category>
  <Domain>person</Domain>
  <Language>hebrew</Language>
  <Definition>father of Kish and Ner, grandfather
```





# Proper Names in the Bible

- About 5,000 different names
- May acquire host language morphology
- Often undergo phonemic transformation
- May carry meanings
- Renderings influenced by
  - Model texts and
  - Local culture



# Consistency

- Developing world vernacular languages
  - Less stable orthographies
  - Orthography changes mid-project
  - Often speakers have lower literacy levels
- Consistent renderings for key characters very important





# Finding Proper-Names in a Text

- Minimal instruction from the user (preferably none)
- Must work for any language
- 2 Possibilities:

## 1) Statistical Glossing

May be defeated by complex morphology or insufficient corpus of text.

## 2) Matching phoneme patterns

Harder, but can work with a very small amount a text.



# The Solution: CogNomen

- Treats a name as an ordered set of phonemes
- Starts with a model name (usually from a local LWC)
- Parses model name into phonemes
- Applies any known phoneme transliteration rules
- Searches target text for similar sequences

The screenshot shows the 'PN Editor' window with the following details:

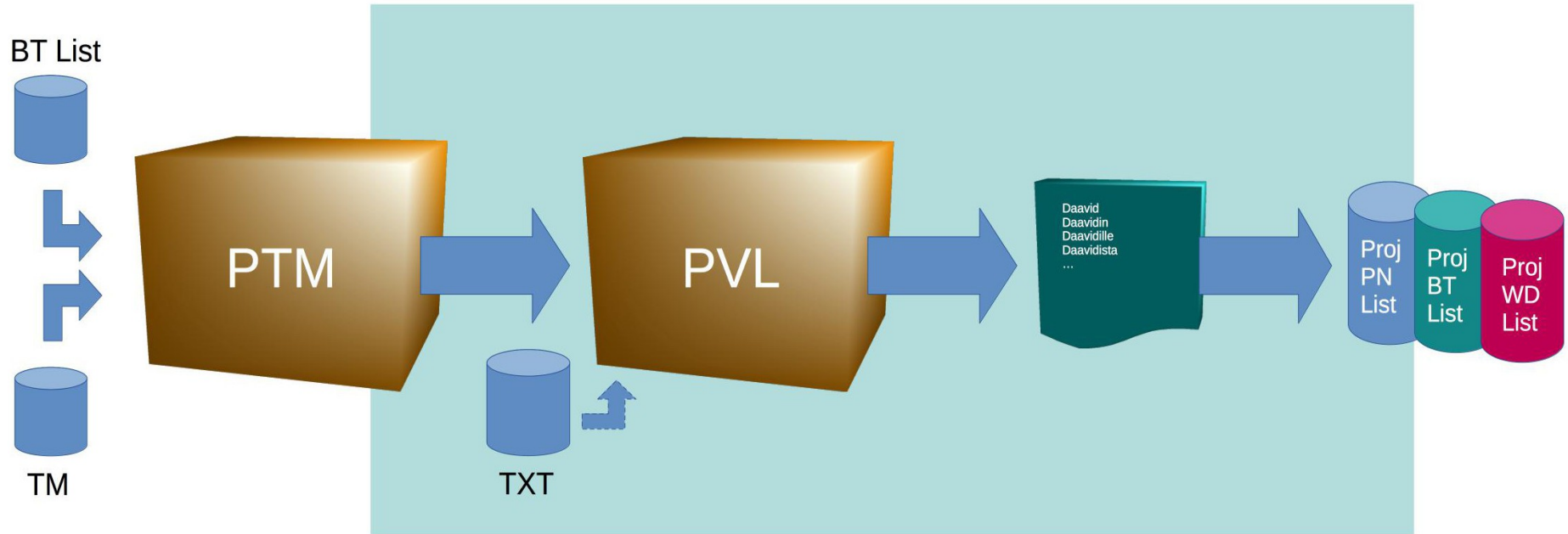
- Forms:** A text box containing 'David'.
- Checklist:** A list of phoneme sequences with checkboxes: daavidin, daavid, daavidille, daavidia, daavidiin, daavidilta, daavidista, daavidilla, daavidkin. All are checked.
- Show match fails:** An unchecked checkbox.
- Morphology:** A dropdown menu showing 'Morphology' and a text box containing 'daavid\_{\_in,\_ille,\_ia,\_iin,\_ilta,\_ista,\_illa,\_kin}'.
- Comments:** A text box containing '0.991'.
- Base Term:** A panel with fields: Term (TJJ), Transliteration (dāwid), Category (PN), Domain (person), Language (hebrew), Definition (son of Jesse; king of Israel), Gloss (David), and Links.
- References:** A scrollable list of biblical references: RUT 4:17, RUT 4:22, 1SA 16:13, 1SA 16:19, 1SA 16:20, 1SA 16:21, 1SA 16:22, 1SA 16:23, 1SA 17:12, 1SA 17:14, 1SA 17:15, 1SA 17:17, 1SA 17:20, 1SA 17:22, 1SA 17:23, 1SA 17:26, 1SA 17:28, 1SA 17:29, 1SA 17:31.
- Buttons:** 'Cancel' and 'Apply' buttons at the bottom right.

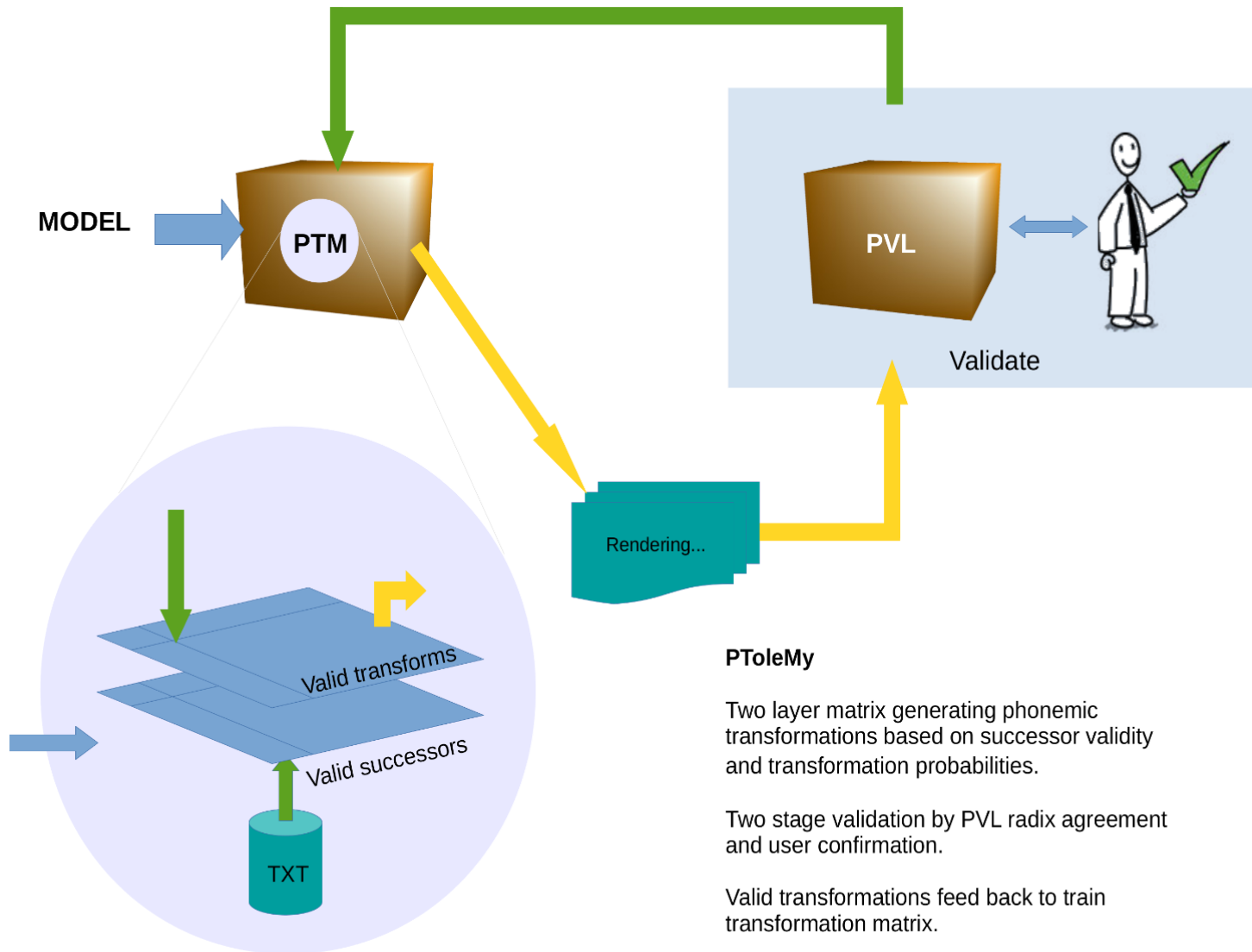
# Two sub-processes

- PToleMy (PTM)
  - Transforms a model name into a form likely to be found in the target text
    - Orthography shift
    - Character clusters
    - etc...
- Percival (PVL)
  - Discontinuous pattern matcher
  - Assesses how close a possible solution is to the target
  - Checks for interference



# CGN – Process schematic:





# Example: Abraham / Ibrahimu

- Abraham
  - Parse: A.b.r.a.h.a.m
  - Get target text
  - Generate word list
  - Test for similar patterns:
- Ibrahimu
  - Parse: I.b.r.a.h.i.m.u
  - Find longest common sequence shared by both
    - `_.b.r.a.h._.m._`
  - Check for 'interference'





# How it works:

- Where:
  - $V$  = match value
  - $dS$  = length of sequence
  - $\theta$  = maximum proximity threshold
  - $d_1$  = the greater of the distances between the two x and the two y coordinates
  - $d_2$  = the lesser of the distances between the two x and the two y coordinates

$$v = \prod_{i=1}^{|dS|-1} \theta - \left( d_1 + \left( \frac{1}{\theta} \cdot d_2 \right) \right)$$



# Example Output - David

- Finnish

- Daavid
- Daavidille
- Daavidia
- Daavidista
- Daavidiin
- Daavidilta
- Daavidin
- Daavidilla
- Daavidin

- Daavid\_ :  
\_ille,  
\_ia,  
\_ista,  
\_iin,  
\_ilta,  
\_kin,  
\_illa,  
\_in

- Bi-Tojol

(Mexico-Mayan)

- David
- Dabid
- Dabidi
- davidi
- Davidi
- Dvidi
- Davida

- D\_id\_ :  
\_av\_i,  
\_av\_a,  
\_v\_i,  
\_ab\_  
\_ab\_i



# Typical accuracy

- Book of Amos
  - 9 chapters
  - 16 personal names
    - 4 languages
      - Finnish
      - Swahili
      - Spanish
      - Indonesian
  - Results:
    - PVL alone 84%
    - PVL + PTM 95%
- Finnish
  - PVL alone 12/16 found
  - PVL + PTM 15/16
    - Mizraim – Egypt
- Swahili
  - PVL alone 14/16 found
  - PVL + PTM 16/16
- Spanish
  - PVL alone 15/16
  - PVL + PTM 15/16
    - Mizraim – Egypt
- Indonesian
  - PVL alone 13/16
  - PVL + PTM 14/16
    - Mizraim – Egypt, Joseph - Yusuf



# Limitations

- Linguistic:
  - Translated names not found
  - Difficult to get valid results with very short names - 'Ai' (עײ)
- Computing
  - Highly compute intensive
  - Matching between element sets with more than 18 members can take a few seconds



# Benefits

- Sits in the background as translators work identifying names as they are entered into the text
- Identified names can be segregated from morphology analysers and/or morphology templates can be used to improve other systems
- Lists cognates together (helps trap spelling errors)
- Creating lists of names for a text becomes very easy



# Wider Application 1

- Proper names are not unique in not typically being translated. Technical vocabulary may also be transliterated between languages.
- CogNomen can automatically build a technical index for a document in any language, based on identifying words which share key term patterns.

**Surprised by the Voice of God**

378	understanding Scripture, 371-72	facilitating prophetic ministry, 176-89
	How God speaks through the Bible, 97-13	in public settings, 186-89
	through experience, 114-29	Prophetic words, 65, 69-77, 176-77, 184, 359
	through natural means, 142-56	how to give, 191-96
	through supernatural means, 180-41	and prayer, 194-95
	Howie, John, 79	in small groups, 199-200
	Humility, 317-20	Prophetic pitfalls, 204-16
	Jesus	Prophecy, 68-69, 302-64, 369
	key to life, 123-24	authority of, 181-89
	limits of his humanity, 43	discerning the present and future, 178
	reliance on the Holy Spirit, 44, 46	Reformation prophets, 69-78
	source of his power, 43	Pullinger, Jackie, 324
	teacher of the heart, 262-64	
	Johnson, Tim, 321-22	Raberg, Jean, John, and Jeanelle, 343-57
	Joyner, Rick, 170, 188	Robison, James, 36
	Kinross Theory, 360-61	Rutherford, Samuel, 84-86
	Knox, John, 70-73, 85	Schaeffer, Francis and Edith, 130-31, 133, 368
	Lase, W. L., 360	Spurgeon, Charles, 89-91
	Law, William, 264	Stallings, Gene, 139
	Lewis, C. S., 82, 107, 115, 175, 236, 238, 268	Supernatural ministry, 368-69
	Lindemann, Allan and Pat, 354-57	conspiracy against, 79-93
	Lord, Peter, 155	ten Boom, Corrie, 87-88
	Marshall, Peter, 367-68	Theology
	Magic, unbelief through, 289-94	God's moral will and, 271-74
	McCavin, William, 79-81	unbelief through, 270-88
	Munzing, Dan & Robin, 91-93	
	Peden, Alexander, 76-78	Walker, Patrick, 81, 365
	Pharisees, 28, 124-25, 239-40, 243, 247-48, 261, 292	Wattinger, Myra, 33-36
	Potter, Don and Christine, 301-2	White, Fowler, 280-83, 372-74
	Pride, 235-50	Whitefield, George, 82
	Prophecy	Why God doesn't speak, 235-88
	can everyone prophesy? 179-80	Willard, Dallas, 27
	discerning the message, 180-81, 196-99	Williams, Carl and Jesse, 349
		Wimber, John, 21, 202, 213, 297-98
		Wishart, George, 70-72, 85

**Scripture Index**

Genesis	110	Judges	6:15	218
1:1	26	7:13	223	223
1:17	336	21:25	294	294
1:27	223	1 Samuel		
2:1-7	217	3:1	43, 345	345
2:3	274	3:7	124	124
2:13	223	3:10	354	354
2:14	223	9:18	178	178
2:17	223	9:20	179	179
4:1-7	220	10:1-9	10:25	225
4:32	220	10:19	225	225
		11:16	304	304
Exodus	319	15:17	294	294
1:11	132	15:23	304	304
3:9	131	15:27-28	178	178
3:18-20	45	18:6-9	225	225
24	45	18:10-11	225	225
24:6	358	28:6-25	294	294
33:11		1 Kings		
Leviticus	508	3:5-15	227	227
11:1-23	110	3:7	228	228
20:18		13:1-32	324	324
Numbers	318	13:1	185	185
12:4-8	132, 318	13:9	185	185
12:6	145, 219	13:18	185	185
16:16	110	16:3	236	236
Deuteronomy		19:9	25, 226	226
4:15-20	131	19:33	207	207
4:29	131	22:4-28		
4:33	131	2 Chronicles		
4:35	131	33:12-13	167	167
4:36	131	20:20	318	318
5:12-29	131	33:12-13		
9:24	102	Nehemiah	150-54	150-54
12:31	328	7:5		
13:1-5	108	Job		
14:3-20	69	7:14	224	224
18:15-22		1:5	224	224
Job	101-102	3:11	224	224
1:5	102	3:14-18	224	224
1:6	102	3:24	224	224
1:7-8	102			
1:7	102			
1:8	102			
1:9	102			



## Wider Application 2

- The process which drives CogNomen was described to this conference in 2012. We call it Percival. Percival is an all-purpose non-contiguous pattern matcher. As such it can find discontinuous patterns in many different contexts including:
  - Automatic analysis of complex, non-contiguous morphologies (such as those exhibited by Semitic languages) for any language.
  - Automatic recognition of syntax patterns in any language. (This work is at an early stage but results are promising).



# United Bible Societies (UBS)

## Glossing Technologies Project

Machine Assisted Translation  
For the Bible Translation community

jonriding@biblesocieties.org  
nboulton@biblesocieties.org  
<http://ubsicap.github.io/mat/>  
<http://www.biblesociety.org/>

