NVIVO Keyness Analysis With Example of Matrix Results

1

Wendy Olsen 2015

Gender Norms Project
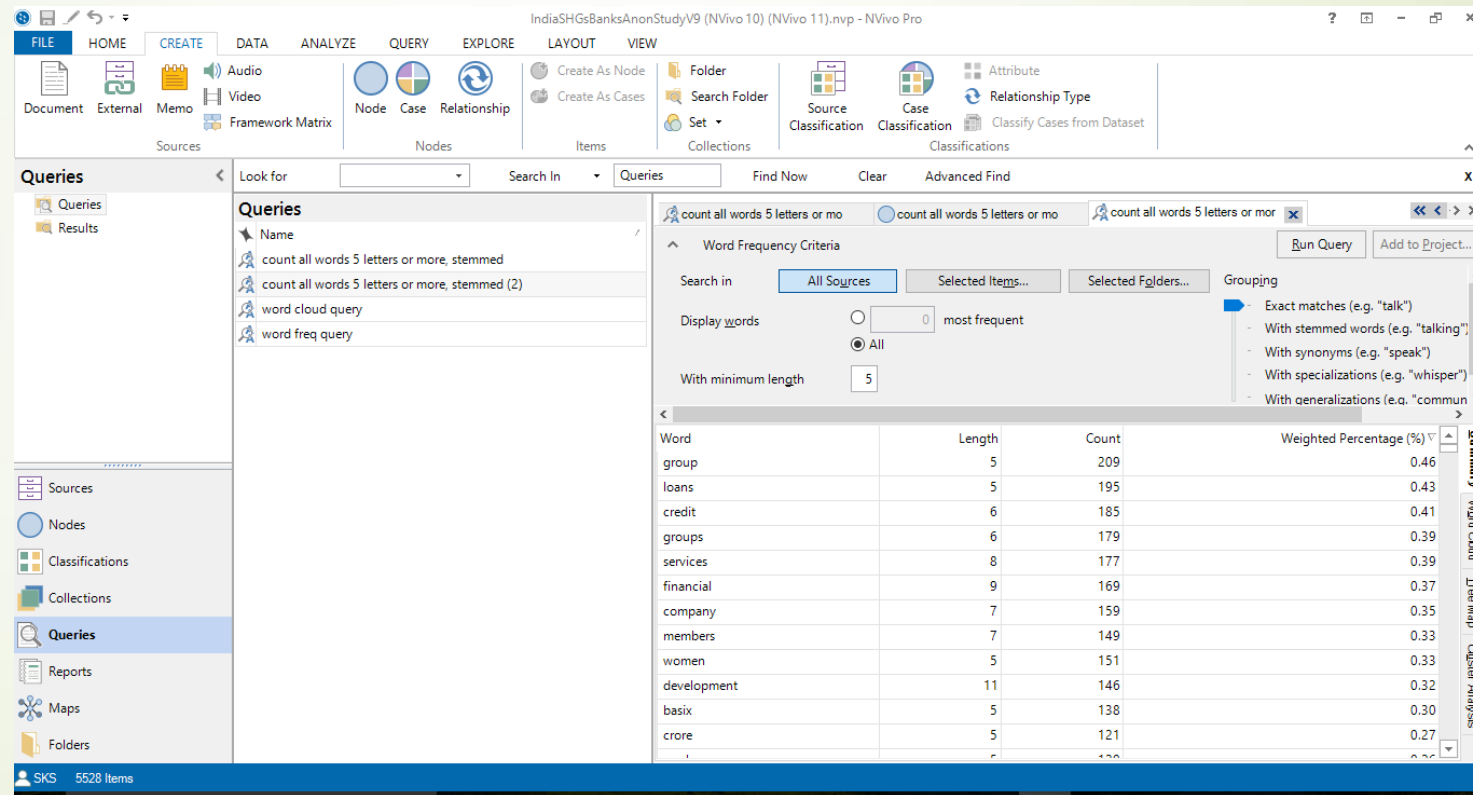
Plan of the work – Proposed. Acknowledging the work of Maheen Sultan, Simeen Mahmud, Mohammed Saiful, Kabita Chowdhury, Sahida Khondaker, Bayazid Hasan, and Sohela Nazneen on the interview side

1. **Pool all the transcripts**

2. **Find out the keyness of words**

3. **Code up the concordances**

4. **Group the words into discourses**

5. **Interpret selected discourses only**

6. **Treat each one of those very carefully: the** dominant discourse **must be discerned, then the** marginalised, deviant and innovative (intertextual**) ones.**

7. **Trace key arguments through these.**

Word Count Query in NVIVO Ranks Words, Giving Word Count Inside the Project

Minimum word length 5
Arbitrary, but cutoff won't matter later

Keyness: the relative prevalence of a key word in a text

- Comparative text is the whole British National Corpus

- Example of the key words and their keyness is shown on the next slide.

- Toure and corpus linguistics offers a dispassionate way to use all the Semi Structured Interview data.

- We must still use our professional judgment to code some additional themes.

- Touri, M., and N. Koteyko (2014) "Using Corpus Linguistic Software in the Extraction of News Frames: Towards a dynamic process of frame analysis in journalistic texts", *International Journal of Social Research Methodology*, DOI: 10.1080/13645579.2014.929878

Odds Ratio

- Take the word's percentage of the words in all the SSI texts.

- Divide it by

- The word's percentage of the count of all those same words in the BNC.
- (Unused words won't matter to the relative prevalence)

- An odds ratio is the relative odds of seeing this word appear in our texts.

Key Words in First SSI

- Top key words by keyness:

- Key words near the 300 rank level

brickfields
laws'
purdah'
coops
passbook
betel
mindset
parishad
stipends
negatively
sons'
educate
workloads

**Interesting,
not restricted in usage to any
one discourse,**

Important in several

calculate
cleaned
oldest
elites
courtyard
scissors
float
meals
earthquake
household
beneficiary
777

Key words at very low rank

important
every
first
group
although
great
course
world
might
still
right

**Anodyne, uninteresting,
non-specific to any
particular discourse**

Excel Spreadsheet – Lowest Keyness

business	8	1	0%	35512	1%	0.13
often	5	1	0%	37138	1%	0.13
later	5	1	0%	38318	1%	0.12
important	9	1	0%	38686	1%	0.12
every	5	1	0%	39090	1%	0.12
first	5	3	0%	120675	2%	0.12
group	5	1	0%	41486	1%	0.11
although	8	1	0%	42736	1%	0.11
great	5	1	0%	47323	1%	0.10
course	6	1	0%	48987	1%	0.10
world	5	1	0%	58170	1%	0.08
might	5	1	0%	59388	1%	0.08
still	5	1	0%	71874	1%	0.07
right	5	1	0%	90541	2%	0.05
	396	1067	100%	5055509	100%	

Excel Spreadsheet – Middle Keyness

difference	10	4	0%	11167	0%	1.70
spent	5	4	0%	11441	0%	1.66
wants	5	3	0%	8716	0%	1.63
medicine	8	1	0%	2915	0%	1.63
amounts	7	1	0%	2995	0%	1.58
speak	5	3	0%	9004	0%	1.58
river	5	3	0%	9162	0%	1.55
everywhere	10	1	0%	3085	0%	1.54
difficult	9	7	1%	21599	0%	1.54
tools	5	1	0%	3129	0%	1.51
better	6	11	1%	34726	1%	1.50
feels	5	1	0%	3202	0%	1.48
young	5	10	1%	32327	1%	1.47
accompanied	11	1	0%	3249	0%	1.46
father	6	7	1%	22757	0%	1.46
covering	8	1	0%	3273	0%	1.45
connected	9	1	0%	3296	0%	1.44

Excel Spreadsheet – Highest Keyness

Delemmatised Word	Length	(mention) Count	of mentions Percent	BNC Prevalence	BNC % of BNC	Ratio Odds
brickfields	11	2	0%	2	0%	4738.06
laws'	5	2	0%	3	0%	3158.71
purdah'	7	1	0%	2	0%	2369.03
coops	5	2	0%	11	0%	861.47
passbook	8	1	0%	6	0%	789.68
betel	5	3	0%	23	0%	618.01
mindset	7	1	0%	11	0%	430.73
parishad	8	2	0%	25	0%	379.04
stipends	8	2	0%	38	0%	249.37
negatively	10	2	0%	52	0%	182.23
sons'	5	1	0%	28	0%	169.22
educate	7	12	1%	365	0%	155.77
workloads	9	1	0%	43	0%	110.19
rears	5	1	0%	45	0%	105.29
chores	6	6	1%	275	0%	103.38
robbers	7	5	0%	244	0%	97.09
tailoring	9	3	0%	147	0%	96.70
dhaka	5	1	0%	55	0%	86.15

Notes on scale of project

- One SSI (Sayera pseudonym)
- We had 673 words of length 5 words or over, of which
- Some were based on the same stem (lemmas). After adjusting,
- We had 396 key words. We did the delemmatisation manually.
- Actually NVIVO can do that.
- For the whole of a south Indian database in NVIVO, there were:
- for example 'scheme' appears 40 times
  - And 'providing' 39, 'provided' 42 times.
- In the unstemmed or destemmed list, which NVIVO did, we got
  - 'Scheme' appeared 40 times, and 'providing' 147 times
  - So we can use NVIVO to reduce our workload considerably in the BNC.

Notes on Scale of Another Project

- Unstemmed word count for SHG project (NVIVO does Query):

- 5529 words of length 5+ letters.

- 3710 word roots of length 5+ letters

- The delemmatisation reduced the length of our key word list by one-third.

Our Project in Bangladesh

1 interview	673 raw words5+	396	By hand
1 interview	459 raw words5+	400	By NVIVO
11 interviews	1666 words	1249 once stemmed	By NVIVO
32 interviews	2798 words	2066 once stemmed	By NVIVO

Using the Words with Highest Keyness

- We set a cutoff level for keyness (the odds ratio) e.g. 4, or 9.

- This means it is 4 times more likely (or more) that we'd see this word rather than the words not chosen, among all the words of length 5 letters or more.

- Obviously 9 times more would mean a word is even more central to the themes of the interview.

- The interviewer affects this keyness a lot, and their words are included in our word counts. That is not necessarily the case, but in this project, it reflects our clear Research Question.

Box 6: Sorted Keyness Table Showing Cutoff Point

Keyword	Number of Appearances	Prevalence in Text	Keyword Number of Appearances in British National Corpus	Prevalence in English (divide first prevalence by a large number)	Keyness (Col. 3 / Col. 5)
Cheat	1	1.000	404	0.003	301.26
Property	8	8.000	12506	0.103	77.86
right	5	5.000	7854	0.065	77.48
lady	4	4.000	9287	0.076	52.42
tenant	1	1.000	2574	0.021	47.28
agricultural	1	1.000	4043	0.033	30.10
Struggle	1	1.000	4329	0.036	28.12
Connection	1	1.000	4984	0.041	24.42
Thousands	1	1.000	5352	0.044	22.74
Husband	2	2.000	10738	0.088	22.67

- Here the South Indian extract is analysed in terms of its key discourses:
- 1) a discourse of property and inheritance;
- 2) a discourse about work of women on land
- 3) a discourse about kin and non-kin roles

The Research Question for Qualitative Part of Gender Norms Project

- RQ1 does women's work fit their attitudes about women's work, or are they coerced by circumstances into working? Or blocked from working when they would like to?

- RQ2 If a woman's role include work that is market work, does a man's activity basket adjust to allow for that, and does their two sets of activities 'fit' well together?

- RQ3 Do the social norms prevalent in a particular area/region influence who works, and if so, does it block the helpful influence of NGO and antipoverty programmes in such an area?

- RQ4 is it a public patriarchy that says women in the rural areas must not work? Is it patriarchal or damaging to suggest that it lowers a household's status for a woman to do work outside the home?

The Same RQ as For the Quants Activities in Each of Papers 1-4

- This is known as tracing a theme, developing an argument, or "following a thread". See for example O'Cathain, Murphy and Nicholl, 2010, BMJ, pages 147-148, and the works by Morgan that she cites there.

- We need to read up some of the health research discussion on triangulation so that we can defend our methods well. Key authors say:

  - Morgan 1998 Qual Health Research *it can be done*. 8: 362-76.

  - O'Cathain et al., *teamwork can work, but respect must be shown to the qual researchers for integration to occur;* Qual Health Research 2008, 18: 1574-85.

  - O'Cathain, et al., J Mix Methods Res 2007: 1: 147-163, *The 'yield' from mixed methods depends on the publication strategy!* **Love the farming metaphor**

Set up a text search query to collect the concordance of 3 sentences each

- For example using the first semi-structured interview. S-----Interview 1.
- (Use Pseudonyms whenever we are in public. Mask the face.)

Reminder of the objectives of the qual part of our work

- To discern contrasts and tensions

- To find out where there is a dominant discourse

- Also to find where is the marginalised, deviant, or competing discourse

- We have to be aware of intertextuality at the same time

- Fairclough and many other writers guide this interpretive work.

Buzz Question   / Exercise

Box 7: Discourses and Keywords

Keyword	Discourse, meaning, discussion
Cheat	1. In the **discourse of property inheritance**, t... rights of recipients and the debate over chea... makes reference to law as well as family connections.
Property	
Thousands	2. A **discourse of family obligations** states th... family members should be loyal to their pare... and lineage. Resistance, desertion, and exceptions are discussed at length. These discussions act partly to discipline those now living in families.
…and right, lady, connection	
tenant	3 In the **discourse of working** on the land for arable farming, the hard work is mentioned v... pride. A quote illustrating that other forms of work are also admired, a man speaks admirin... of a good woman doing "clay work, open wel... work and tank work"
agricultural	
Struggle	1. *(sic)* In the **discourse of family obligations**, inheritance exceptions are hotly debated, and some go into court, and last for decades. Both families support their strugglin... parties who champion their rights, obtained
Connection	

- Sample of3 discourses for South India.

- Can we examine the one numbered 3 here in relation tour example of a man discussing a women doing "hard work", tank work, etc

agricultural 21	3 In the discourse of working on the land for arable farming, the hard work is mentioned with pride. In a quote illustrating that other forms of work are also admired, a man speaks admiringly of a good woman doing "clay ==work==, open well ==work== and tank ==work=="= RED QUOTE
Struggle	1. (sic) In the discourse of family obligations, inheritance exceptions are hotly debated, and some go into court, and last for decades. Both families support their struggling parties who champion their rights, obtained through connection and family lineage, although they are also pitted against other claims of family rights.
Connection	
Right	
Husband	4. Gender roles are often referred to via individual people's family roles. But when a person is speaking with disapproval about a woman, the term 'lady' is used (in translation) without a

Interpret the red quote

a) as morphogenetic

b) as morphostatic (promoting change)

The 3-minute segment – South India

In my father's period, we are three male children. When fourth child was born my mother died. She was our own mother. After her death I was continuing of his agricultural labour work. Day by day the family was broke down. Then my father went to ….After that my father had a relationship with other lady. And he came back with that lady and put a small hut here NO SUCH A GOOD HOUSE BUT SHE WORKED HARD. SHE DID CLAY WORK, OPEN WELL WORK AND TANK WORK ALSO. In my childhood I was feeding village people cows 10 years. SHE BUILT A HOUSE IN THE PLACE OF HUT. And she put us happily at last she got news. Her first husband left her and had a connection with other lady. He was working in a railway station and he died. After having got the news she went there. She had responsibility. [MEANS, SHE WILL INHERIT IN PREFERENCE TO HER FIRST HUSBAND'S SONS AND HIS SECOND COHABITING WIFE] But she did not go there. After that some people came here and took her with them. My father also went with her…. [HER 1st HUSBAND DIED] Actually that lady [2nd COHAB WIFE] had a right, and she also had a right on his property. Both had right on his property. After that some people said to her like this. "See lady! Your husband died. …you had a right on h is property". On that side other lady who had connection with him, she had some children, and…. had also right on his property. Like this they fight and struggled each other. At that time my father spent 20 to 30 thousands for this case. After a long period she got that property. After that all brothers were fought for this property. Later after getting of this property she joined with her brothers and my father was left alone. Like THIS MY FATHER WAS CHEATED.

Morpho-genetic = when we promote and admire women working hard on outside land, with bricks, water (tank), shrimp, animals or trading, we encourage their modern involvement in business, too, and in employment – a form of pride.

Morphostatic OR NOT?

- When we say that the woman is CHEATING this man of some property, we are saying that by LEAVING him and "MAKING" him spend 20,000 Rupees on a court case of "HERS", she is to blame for him WASTING his money,

- As an observer we are at that point PUTTING her in a PASSIVE, yet GUILTY position as a waster and gold-digger We are avoiding the issues of LOVE and COMITMENT where she in fact had had 2 families, 2 sets of commitments, & probably was torn in her heart. This heartless formulation is the PATRIARCHAL formulation. Her stepson sees it this way.

The usual epistemological riddle

The speaker is post-patriarchal and really respects this hard-working woman. =morphogenetic, **Tension is deliberately created against tradition.**	**The speaker is patriarchal but grudgingly gives some approval to her accumulation strategy and her work capability**
The researcher is trying to be post-patriarchal so we wishfully stress the modern aspect, the deviant and pro-woman aspects of the quote, like 'HARD WORK' as a respectful mode of address.	The researcher should really notice that the overall conclusion made by the speaker is the patriarchal one that this woman is To-Be-Criticised-No-Matter-How-Hard-She-Works. Her 'strategies' are not taken seriously, and are undermined by this family member in the long run.

Plan of the work = see diagram in long text.

1. Pool all the transcripts

2. Find out the keyness of words

3. Code up the concordances

4. Group the words into discourses

5. Interpret selected discourses only

6. Treat each one of those very carefully

7. Trace key arguments through them.

Results for two discourses (family talk and money talk) [india 1 and bangla 1 combined] Mentioned within 30 words of each other, in combination.

E : works	A : Tightness node	B : money	C : problems	D : spend	E : works
1 : Family	20	18	11	9	22
2 : children	12	10	6	5	11
3 : daughter	15	6	3	0	13
4 : husband	10	6	8	1	11
5 : mother	11	6	1	4	11

Interpretation of two discourses' intertextuality (family talk and money talk) [india 1 and bangla 1 combined] Mentioned within 30 words of each other, in combination: Discourse analysis.

E : works	A : Tightness node	B : money	C : problems	D : spend	E : works
1 : Family	20	18	11	9	22
2 : children	12	Children are discursively constructed to be instructed, told, shaped. However in relation to money problems they are told to work if poor, or they are the investment destination while parents aim for their future earnings and wellbeing, if not poor.			
3 : daughter	15				
4 : husband	10				
5 : mother	11				

Coding of the Cases by the Classification (Attributes sex, female-headed household, marital status, age and so on)

Features of People Interviewed	Sex	Age Group	Marital Status Now	Head of Household's Sex
Nodes\\Cases\\Bangla S1	Female	Older 51+ Years	Widow, Prev. Married	Female
Nodes\\Cases\\India S1	Female	Young <25	Married	Female

Now we can query how the Cases used the Discourses

- This is our second matrix query. It can also be done as a 'Framework Query' which implies Cases by Themes.

	A : ESRC_S-----_S----Int1.040915	B : Tr_UP_Jaunpur_Rehati
1 : children	11	8
2 : daughter	15	2
3 : husband	10	10
4 : mother	7	4

Conclusions

- By doing the teamwork in stages, the best use can be made of different kinds of expertise – translation, noticing prevalences, interpreting.

- By combining the Keyness stage with a selective interpretation stage, we leave some further interpretation to do later, or for others to do.

- The same thing is done in statistics where models are debated in an ongoing, open-ended way.

- We try to make the interpretation match, complement or contradict the original Research Question and commonly accepted factual statements. IMMN link https://www.facebook.com/groups/438437119631157/

References

- Heale, R., and D. Forbes (2013), "Research Made Simple: Understanding triangulation in research", *Evid Based Nurs* 16:4 98 doi:10.1136/eb-2013-101494.

- Morgan, J., and W.K. Olsen (2011) "Aspiration Problems for the Indian Rural Poor: Research on Self-Help Groups and Micro-Finance", *Capital and Class*, June, 35:2, 189-212, doi: 10.1177/0309816811402646;

- O'Cathain, A., E. Murphy, and J. Nicholl (2010) "Three Techniques for Integrating Data in Mixed Methods Studies", *British Medical Journal*,341: , doi: 10.1136/bmj.c4587

- Olsen, W.K., and J. Morgan (2010) "Institutional Change From Within the Informal Sector in Indian Rural Labour Relations", *International Review of Sociology*, 20:3, 535-553, on "The Regulation of Indian Informal Capitalism", DOI:10.1080/03906701.2010.51190.)

- Touri, M., and N. Koteyko (2014) "Using Corpus Linguistic Software in the Extraction of News Frames: Towards a dynamic process of frame analysis in journalistic texts", *International Journal of Social Research Methodology*, DOI: 10.1080/13645579.2014.929878