

# Shifting discourse-semantics of risk in US newspapers, 1987–2014

Daniel McDonald   Jens Zinn

@interro\_gator

This slideshow is available at: <http://git.io/vBfbw>

22nd November 2015

- Context of the investigation: risk theory
- Data and research questions
- Linguistic approaches to risk
- Our methods and linguistic findings
- Sociological significance of the results

This slideshow is available at: <http://git.io/vBfbw>

I'm presenting work from closely related projects:

- ① Risk words in the NYT, 1963, 1987–2014
- ② Risk words in NYT health articles
- ③ Risk words in six US newspapers, 1987–2014

All investigations involve making longitudinally structured, parsed corpora and looking at how risk words behave.

From previous sociological and linguistic research, we know that:

- Risk as concept is sociologically important
  - ▶ New global risks (Beck, 1992)
  - ▶ Calculative technologies (Dean, 1999)
  - ▶ Individualisation (Beck) and Technologies of the Self (Dean, 1998)
  - ▶ Risk-taking (Luhmann, 1993)
- Risk as lexical item is increasingly frequent in print journalism (Zinn 2011)
- Risk as a lexical item in naturalistic text may behave contrary to expectations (Hamilton, Adolphs, & Nerlich, 2007)

- ① *NYT Annotated Corpus*: 1.8 million articles, 1987–2007 (Sandhaus, 2008)
- ② *ProQuest Newsstand* for NYT 2007–2014
- ③ *ProQuest Newsstand* for five other newspapers, 1987–2014
  - ① *Washington Post*
  - ② *Tampa Bay Times*
  - ③ *USA Today*
  - ④ *Chicago Tribune*
  - ⑤ *Wall Street Journal*

- What are risk words doing in the NYT?
- How has the behaviour of risk words changed in the NYT between 1963 and 2014?
- Can we connect these findings to sociological theories of risk?
- What kinds of tools and methods can we use/develop to do this kind of research?

New kinds of data and tools make it possible to empirically analyse risk language in new ways:

- Digitisation of newspapers means we have large, well-structured datasets
- Automatic annotation of text makes it possible to search for lexical and grammatical features in tandem
- Modern programming languages facilitate:
  - ▶ Automation
  - ▶ Reproducibility
  - ▶ Transparency

Frame semantics: risk as a cognitive schema (Fillmore & Atkins, 1992)

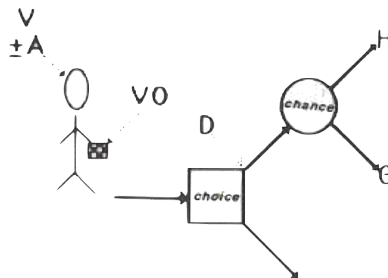
- Conceptualises risk mostly as experiential Process/Event
  - ▶ *What kind of participants and circumstances occur when risk is the Process?*
- Problem: risk often takes less prominent experiential roles
  - ▶ Is the risk frame actually invoked when the word is used?
  - ▶ Example:

*Mr. Tepfer noted that Mr. Douglas, who was in the neighborhood when the body was found and was interviewed by the police at the time, 'preyed on at-risk women, on prostitutes, and he engaged in sex and strangled them to death.'*



# The risk frame

H = Harm, G = Goal,  
D = Deed, VO = Valued Object,  
V = Victim, A = Actor.



Corpus linguistics: risk as token (Hamilton et al., 2007)

- Topics and text-types in which risk tokens appear
- Collocates of risk tokens
- Risk appears a lot in discussions of health
- Use of risk words is different to invented examples

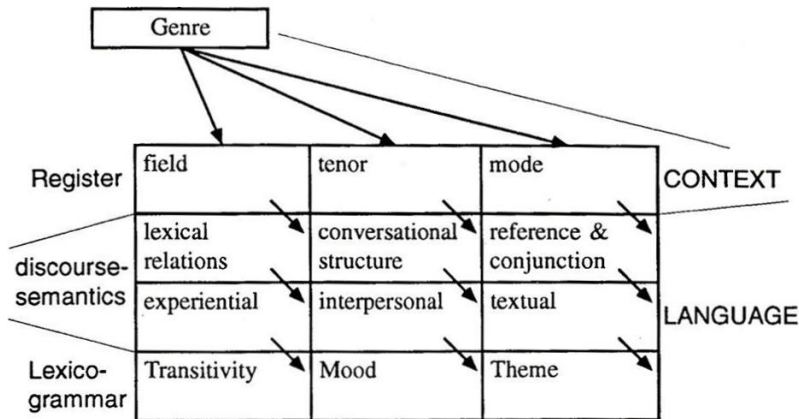
Shortcomings:

- Smaller corpus size, heterogeneity of samples
- No parsing, lemmatisation
- No systematic connection of lexicogrammatical patterns to discourse/meaning

- Get all paragraphs containing \brisk in all 1987–mid 2014 articles
- Annotate/parse the data with full *Stanford CoreNLP* suite
- Develop **corpkit**, a toolkit for manipulating the corpus and communicating results
- Interrogate the corpus according to notions from systemic functional grammar
- Connect to sociological theory

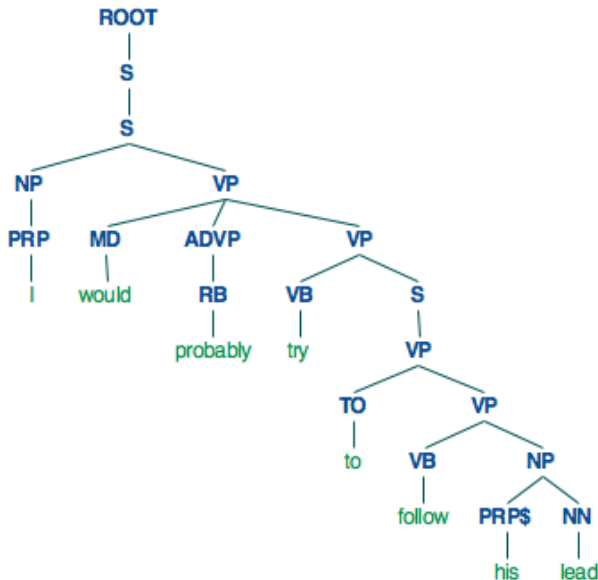
- *Systemic*: focus on relationships between signs in the sign-system
- *Functional*: focus on language as a tool for the performance of functions
  - ① Interpersonal: negotiating relationships
  - ② **Experiential: representing the world**
  - ③ Textual: reflexive organisation into meaningful sequences

In general, print news in a single publication is fairly stable in terms of interpersonal and textual features.

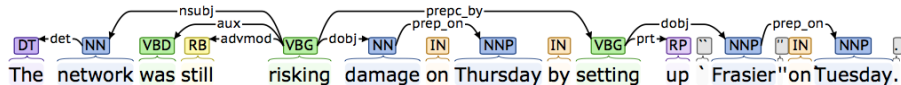


- Focus on the clause as a unit of analysis
- Centre on the *process* (i.e. rightmost verb in VP)
- Processes *select* participants (i.e. arguments of the verb)
- PPs and RBs are typically *circumstances*

# Constituency grammar



# Dependency grammar





# The controversial question

The question: *Can we get systemic functional information from constituency and dependency parses?*

The answer: *Yep, quite a lot.*

How to investigate this huge dataset, and make the investigation transparent/reproducible?

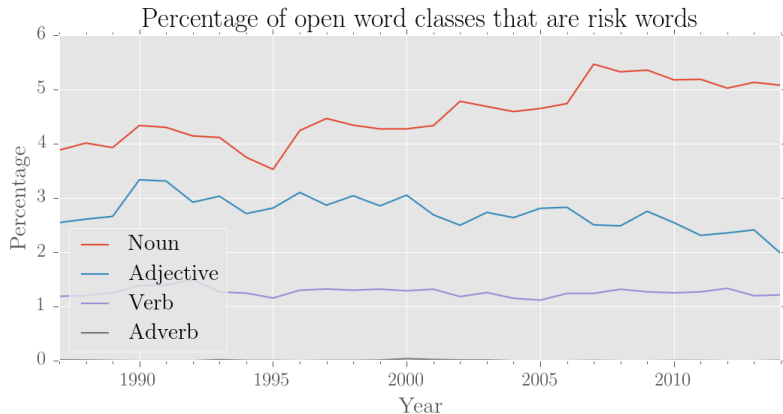
- **corpkit**: a Python module designed for parsed and structured corpora, with some systemic functional awareness
  - ▶ **interrogator()**: search for lexicogrammatical phenomena in each subcorpus, tally results, output Pandas objects
  - ▶ **editor()**: edit results, calculate keyness, linear regression
  - ▶ **plotter()**: visualise via *matplotlib*
  - ▶ **conc()**: concordance via parses
- Scitable, multiprocessing, handles arbitrary data, open-source
- More recently, a GUI, aimed at corpus linguists

```
# import module
>>> from corpkit import *
# set path to data
>>> corpus = 'data/NYT-parsed'
# get pos of risk words, show word class
>>> res = interrogator(corpus, 'words', r'\brisk',
...     show = ['p'], lemmatise = True)
# get relative frequency
>>> rel = editor(res.results, '%', res.totals)
# visualise
>>> plotter('POS of risk words', rel.results)
```

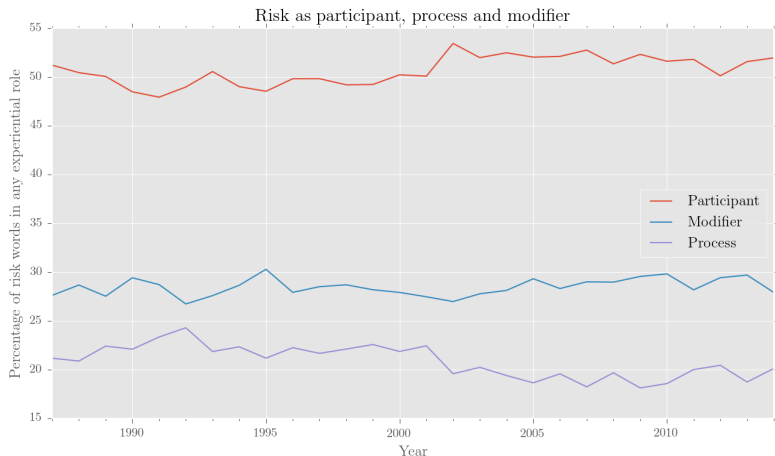
# Initial investigation: NYT

Our first investigation was of the NYT only:

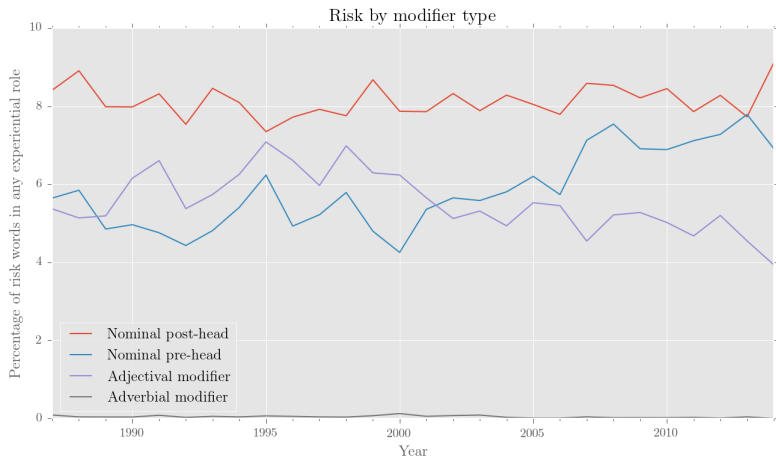
# Output: nominalisation of risk



# Experiential roles of risk words

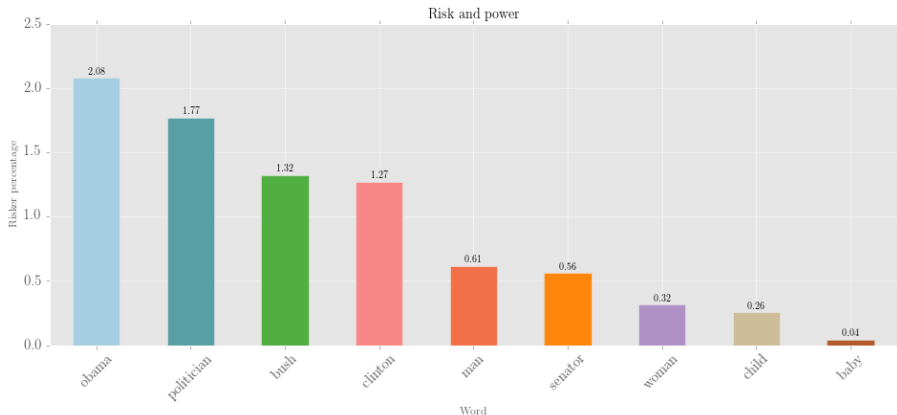


*They risked their life → It was a risk*



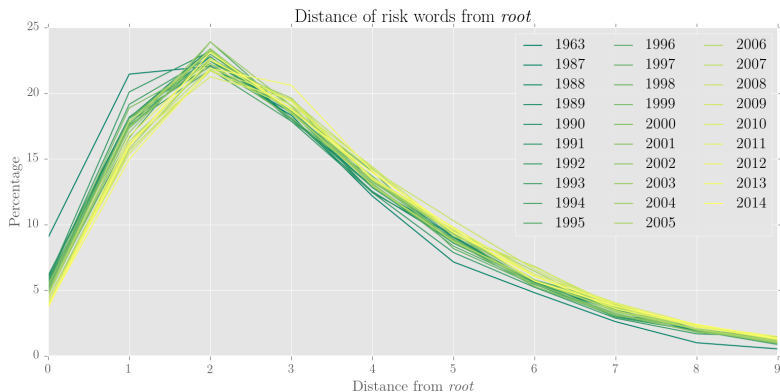
*Risky decision → risk assessment*

# Risk and power





# Distance of risk word from *root*



... looked promising, but seems to be a general phenomenon.

- Nominalisation and *participation*
  - ▶ risking harm → risk assessment
  - ▶ Meaning of risk expanding beyond the *risk frame*
- Risk words becoming more implicit
  - ▶ Routinisation of the management of risk
  - ▶ Risk as increasingly present, but decreasingly debated
- More everyday exposure to risk, but less risking
  - ▶ Neoliberal conceptualisations of agency: institutional expectation to take risk
  - ▶ Reporting of '*the scandal of not being in control*' (Beck, 1992)



- We've only just started interrogating the six newspaper corpus.
- First, we'd like to check if the NYT findings are generalisable to other publications.
- Would love help on dealing with the complexity of the data structure





- Most phenomena generalisable
- Some newspaper specific constructions: *risk appetite* in the WSJ
- Fewer grammatical riskers, but risk characterising more participants and processes
- Hints of influence of newspaper's political position
-

- SFL proves a useful means of dividing up and investigating the behaviour of a given word
- SFL parsing is difficult, as is converting concepts from (esp. formal) grammars
- Difficult SF concepts: rank shift, grammatical metaphor, appraisal, process types (Yan, 2014; Costetchi, 2013; Heyvaert, 2003)
- That said, though theoretical orientations are different, much of the grammar (esp. at group/phrase levels) are similar



Though SFL treats context as embedded in the lexicogrammar of texts, sociological theory can theorise the influence of salient events, people

- Did Chernobyl/Sept. 11 *change* language use in the NYT?

Functional linguistic theory and corpus/computational linguistic provide sociology research with:

- Empiricism
- Reproducibility

Data and tools are available for reuse:

- <https://www.github.com/interrogator/risk>
- <https://www.github.com/interrogator/corpkat>

Findings are presented dynamically in an IPython Notebook:

- <http://git.io/vIM2W>

This slideshow:

- <http://git.io/vBfbw>

- Beck, U. (1992). *Risk society: Towards a new modernity*. Sage.
- Costetchi, E. (2013). Semantic role labelling as SFL transitivity analysis. *ESSLLI Student Session 2013 Preproceedings*, 29.  
Retrieved from <http://stus2013.loriweb.org/files/2013/09/SFL-transitivity-analysis-%25E2%2580%2593-an-alternative-approach-to-semantic-role-labelling.pdf>
- Dean, M. (1998). Risk, calculable and incalculable. *Soziale Welt*, 25–42.
- Dean, M. (1999). *Governmentality: Power and rule in modern society*. SAGE Publications, Inc.
- Fillmore, C. J., & Atkins, B. T. (1992). Toward a frame-based lexicon: The semantics of RISK and its neighbors. *Frames, fields, and contrasts: New essays in semantic and lexical organization*, 103.

- Hamilton, C., Adolphs, S., & Nerlich, B. (2007, March). The meanings of 'risk': a view from corpus linguistics. *Discourse & Society*, 18(2), 163–181.
- Heyvaert, L. (2003). Nominalization as grammatical metaphor: On the need for a radically systemic and metafunctional approach. *status: published*.
- Luhmann, N. (1993). *Risk: A Sociological Theory*. New York: Walter de Gruyter.
- Sandhaus, E. (2008). *The New York Times Annotated Corpus LDC2008T19*. Linguistic Data Consortium.
- Yan, H. (2014). Automatic labelling of transitivity functional roles. *Journal of World Languages*, 1(2), 157–170. Retrieved from <http://www.tandfonline.com/doi/pdf/10.1080/21698252.2014.937563>