

# Discourse-semantics of risk in *The New York Times*, 1963–2014: a corpus linguistic approach

Zinn & McDonald

This slideshow is available at: <http://git.io/vYOM4>

July 2015

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

This slideshow is available at: <http://git.io/vYOM4>

From previous sociological and linguistic research, we know that:

- Risk as concept is sociologically important (Beck, Giddens)
- 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)

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
  - ▶ *NYT Annotated Corpus*: 1.8 million articles, 1987–2007 (Sandhaus, 2008)
- Automatic annotation of text makes it possible to search for lexical and grammatical features in tandem
- Modern programming languages facilitate:
  - ▶ Automation
  - ▶ Reproducibility
  - ▶ Transparency

- The *Risk Society* (Beck, 1992)
  - ▶ Increased centrality of *risk* in daily life in late modernity
  - ▶ *Mega risks*
  - ▶ Individualisation of responsibility for managing risk
- Technologisation: new data and tools for calculating risk and risk factors creates new functions of governments and institutions
- Risk vs. danger (Luhmann, 1989)

We wanted to build on these earlier findings, and take advantage of new technologies:

- 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?

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?

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.'

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

- Topics and text-types in which risk tokens appear
- Collocates of risk tokens (Hamilton et al., 2007)
- Risk appears a lot in discussions of health
- Use of risk words is different to invented examples

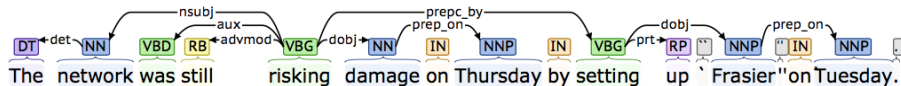
Shortcomings:

- Smaller corpus size, heterogeneity of samples
- No parsing, lemmatisation
- No means of connecting lexicogrammar to meaning



- Get all paragraphs containing *risk* in all 1987–mid 2014 editions of the NYT:
  - ▶ 153,828,656 words
  - ▶ 149,504 articles
  - ▶ 240,08 risk words
- Annotate/parse the data for lemmata, constituency, dependency (not SFL!)
- Develop toolkit for manipulating the corpus and communicating results
  - ▶ <https://www.github.com/interrogator/corpkit>
- Interrogate the corpus
- Connect to sociological theory

Q: Can we extract SF grammatical features from this?



A: Yes, quite often.

```
# import module and define corpus path
from corpkit import *
nyt = 'data/nyt/years'

# count every pos
baseline = interrogator(nyt, 'pos', 'any', lemmatise = True)

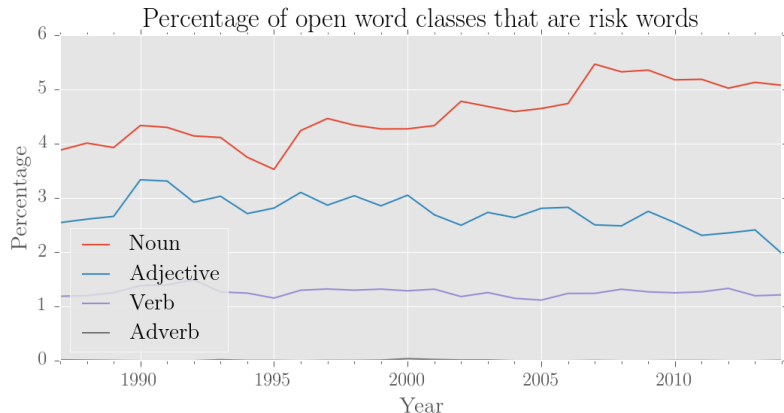
# count pos for risk words
riskp = interrogator(nyt, 'pos', r'__ < /(?(i)\brisk/)', lemmatise = True)

# list open word classes
open_words = ['Noun', 'Verb', 'Adjective', 'Adverb']

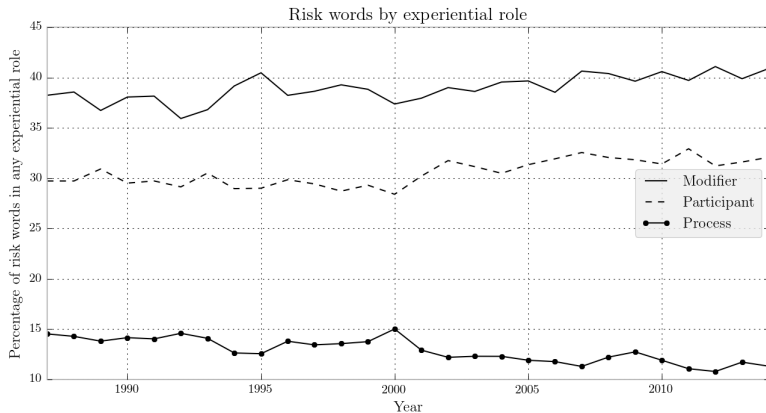
# get relative frequencies of open word classes, skip 1963
maths_done = editor(riskp.results, '%', baseline.results,
    sort_by = 'total', just_entries = open_words,
    skip_subcorpora = [1963])

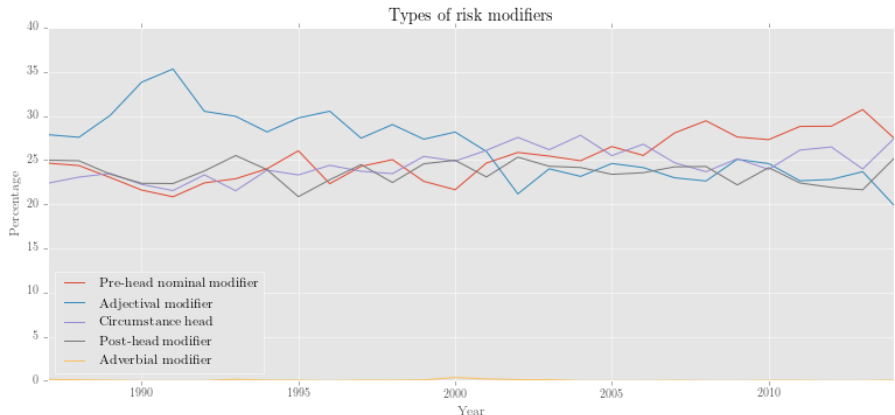
# plot
plotter('Percentage of open word classes that are risk words',
    maths_done.results, legend_pos = 'lower left')
```

# Output: nominalisation of risk

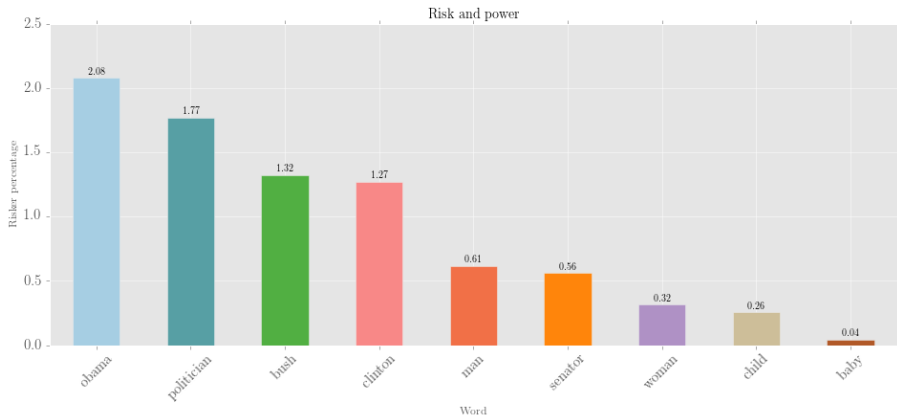


# Experiential roles of risk words

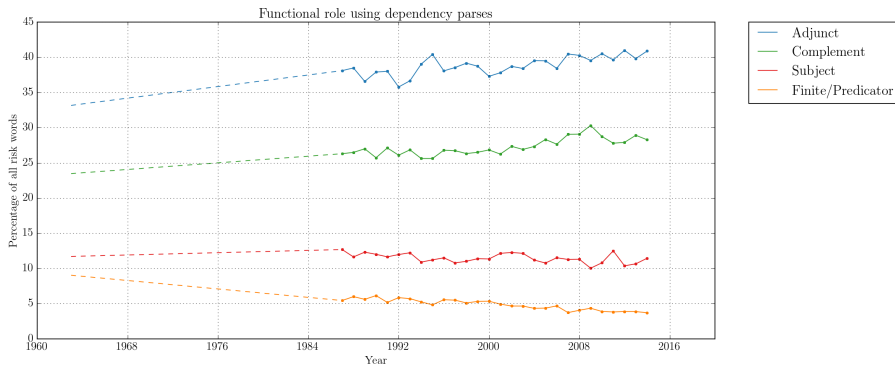




# Risk and power



# Mood role of risk words





- Nominalisation and participantification
- Risk words becoming more implicit
- More everyday exposure to risk, but less risking
- Complicated constellations of participants and circumstances in risk processes

- 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
- That said, though theoretical orientations are different, much of the grammar (esp. at group/phrase levels) are similar

There are points of convergence, as well as disparities between, our linguistic findings and influential sociological theories:

- 
- 
- 
-

- Further exploration of risk as per SFG: process types, mood features, thematic metafunction
- New datasets and comparative analyses
- Expanding our focus to related terms: *danger*, *(in)security*, etc.

Data and tools are available for reuse:

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

Findings are presented dynamically in an IPython Notebook:

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

This slideshow:

- <http://git.io/vYOM4>

- Beck, U. (1992). *Risk society: Towards a new modernity* (Vol. 17). Sage.
- 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. doi: 10.1177/0957926507073374
- Luhmann, N. (1989). *Ecological communication*. University of Chicago Press.
- Sandhaus, E. (2008). *The New York Times Annotated Corpus LDC2008T19*. Linguistic Data Consortium.