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/vY0M4`

July 2015

Presentation overview

THE UNIVERSITY OF MELBOURNE

- Context of our investigation: risk theory
- Our 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/vY0M4`

Context of our investigation: risk theory

From previous sociological and linguistic research, we know that:

- Risk as concept is sociologically important
  - New global risks (Beck, 1992)
  - Calculative technologies (Dean, 1998, 2010)
  - Individualisation (Beck) and Technologies of the Self (Foucault, Martin, Gutman, & Hutton, 1988)
  - 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)

Data and research questions

1. *NYT Annotated Corpus*: 1.8 million articles, 1987–2007 (Sandhaus, 2008)
2. *ProQuest Newsstand* for articles containing a risk word between 2007–2014

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?

New methodologies

THE UNIVERSITY OF MELBOURNE

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 semantic approach

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.'

Corpus linguistic approach

THE UNIVERSITY OF MELBOURNE

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

Our methods

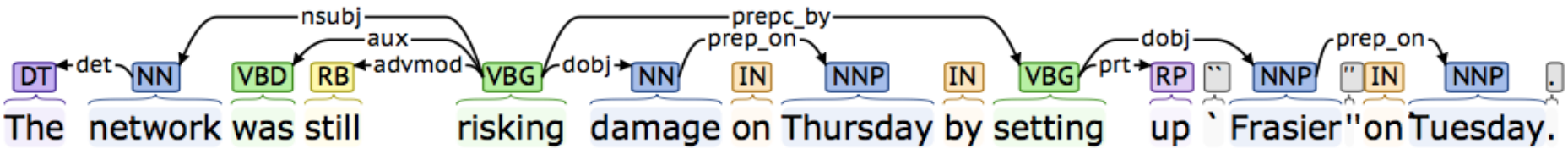
THE UNIVERSITY OF MELBOURNE

- 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 `corpkit`, a toolkit for manipulating the corpus and communicating results
  - `https://www.github.com/interrogator/corpkit`
- Interrogate the corpus
- Connect to sociological theory

Dependency parsing

THE UNIVERSITY OF MELBOURNE

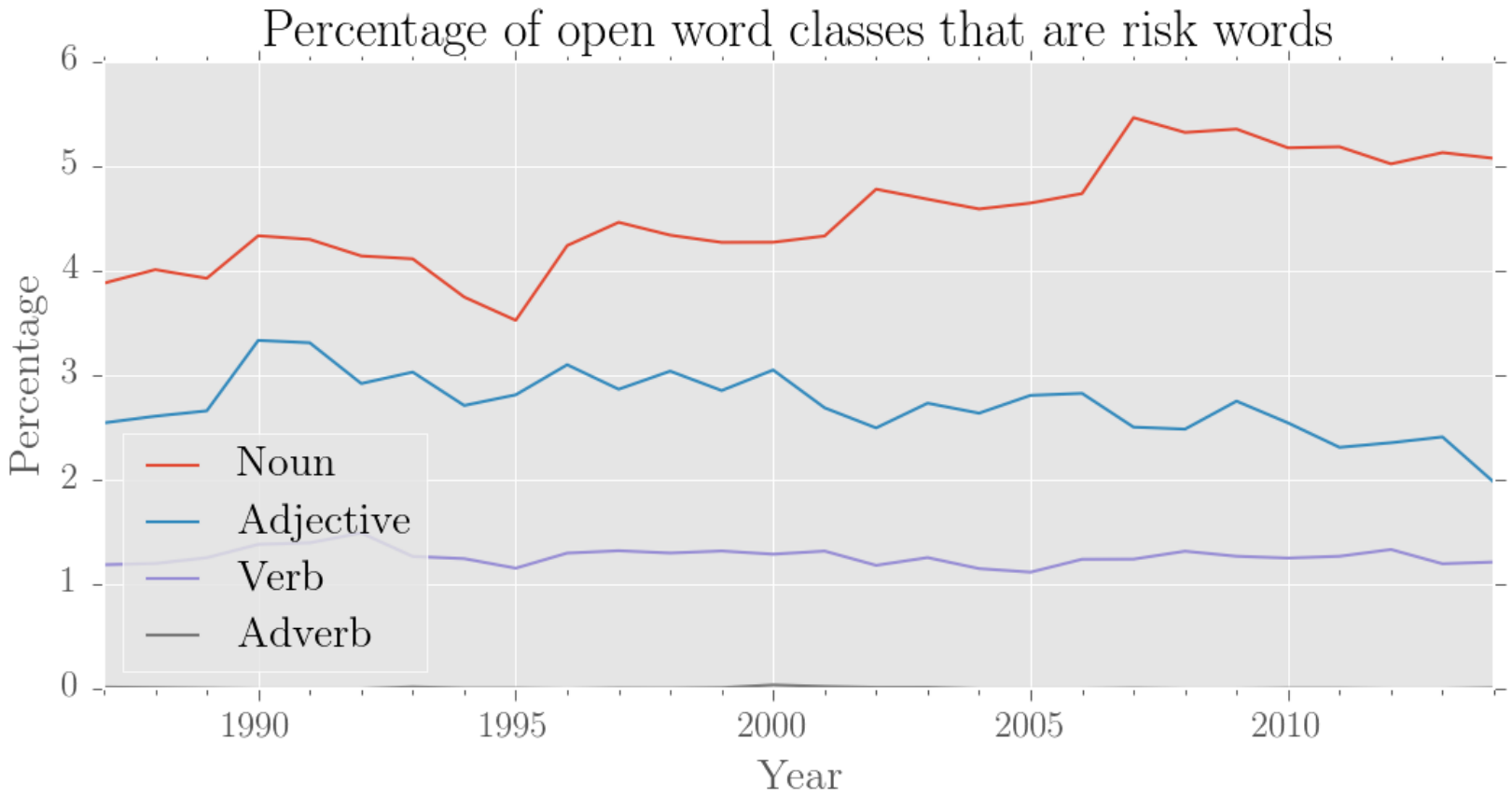
Q: Can we extract SF grammatical features from this?



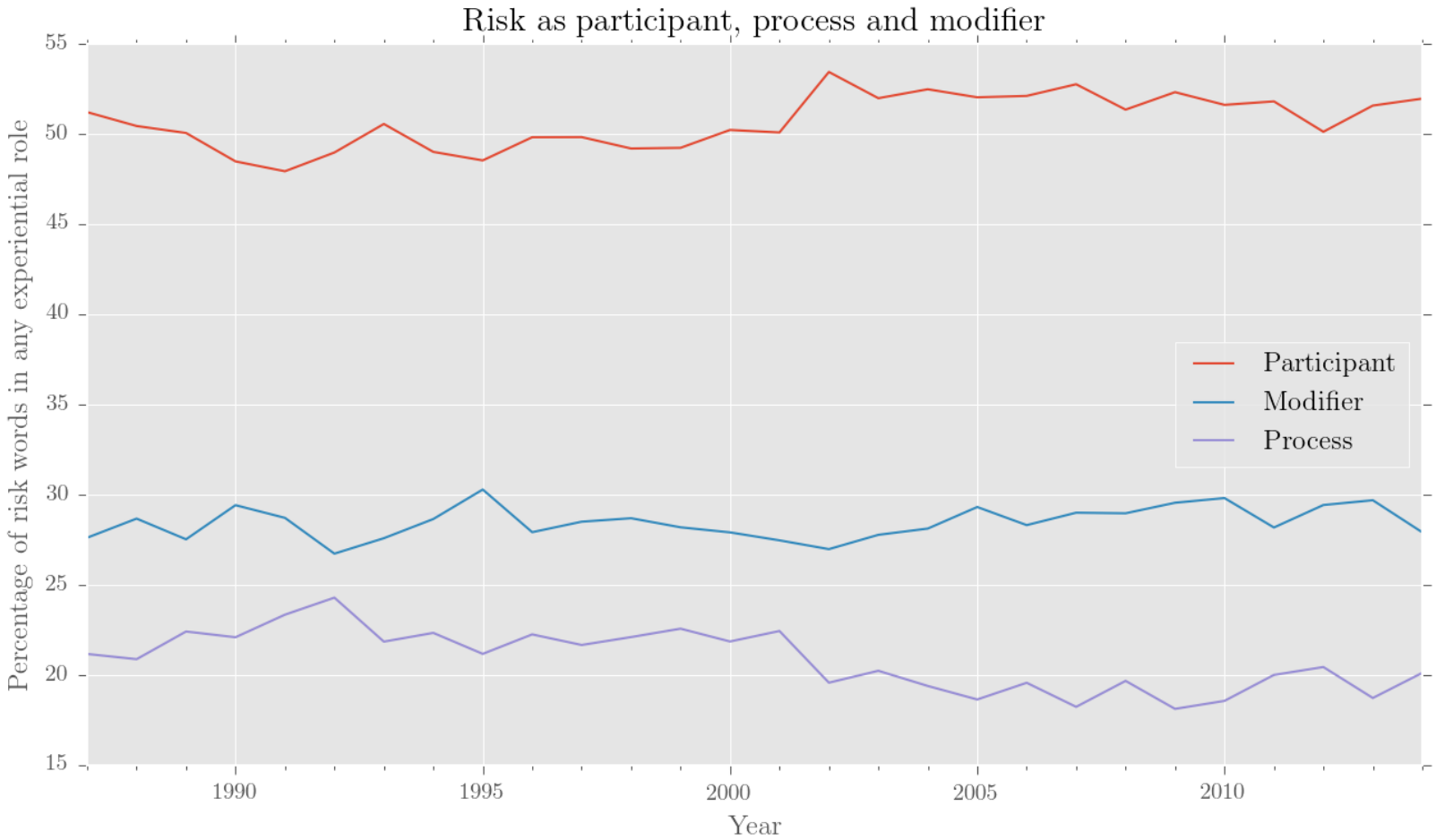
A: Yes, quite often.

```
# 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

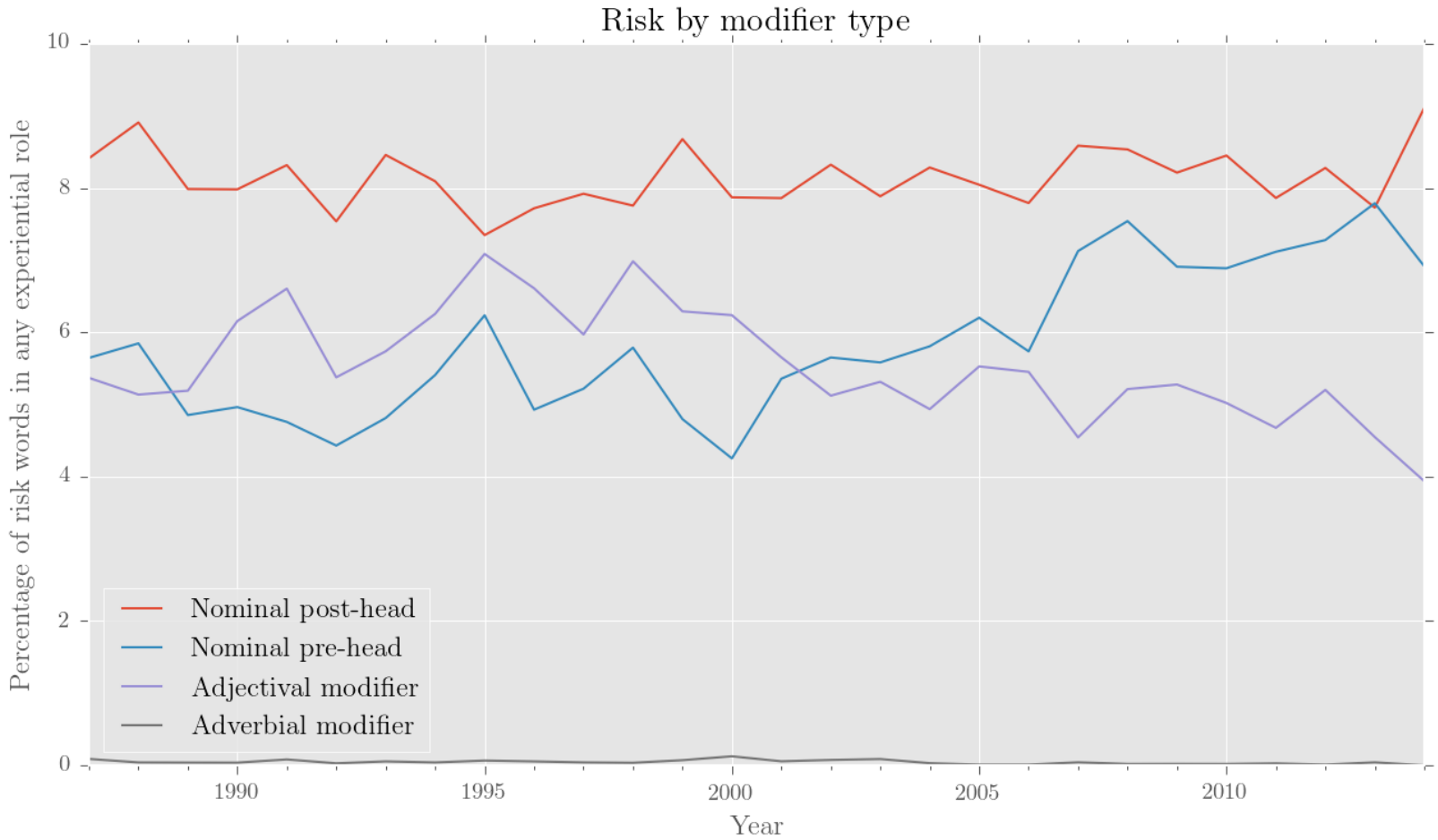


Experiential roles of risk words



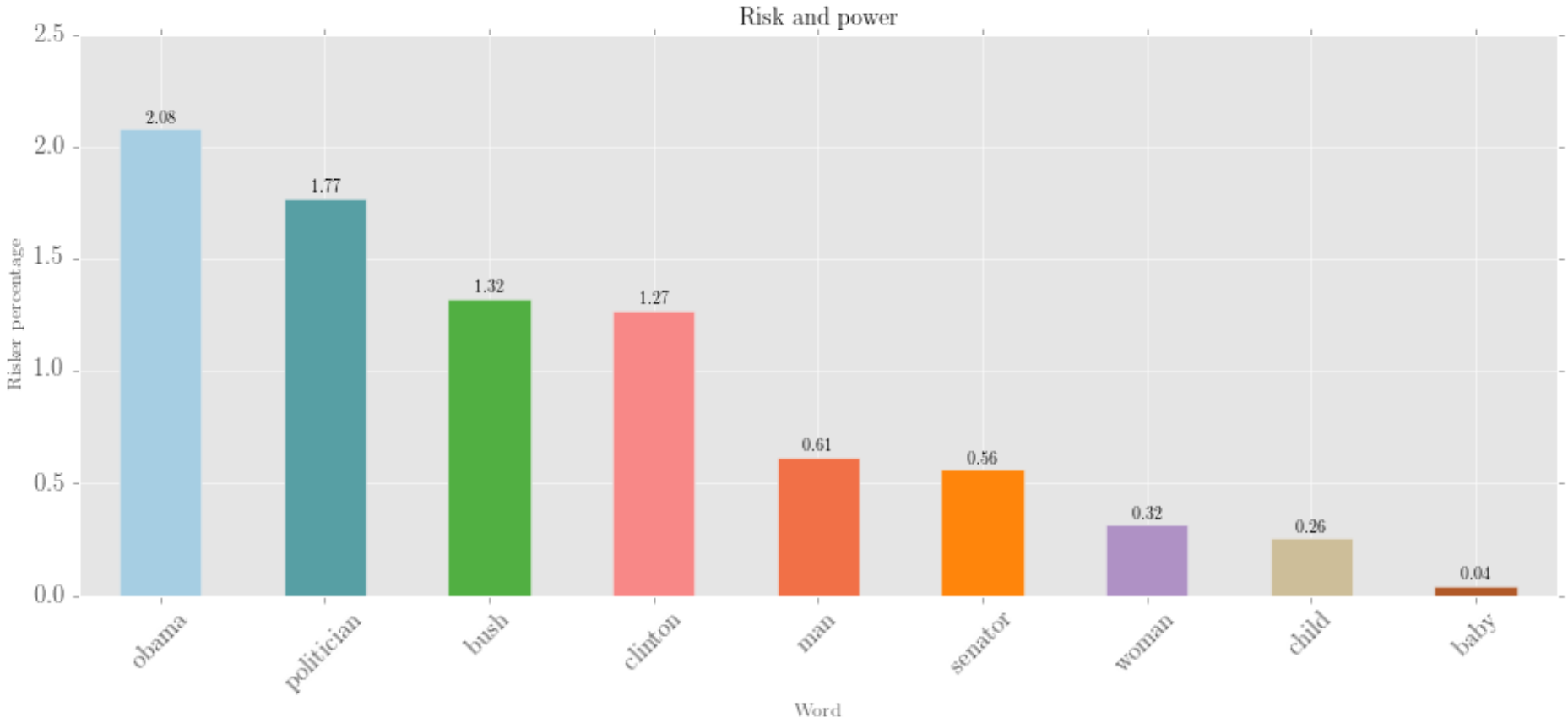
They risked their life → It was a risk

Risk as modifier

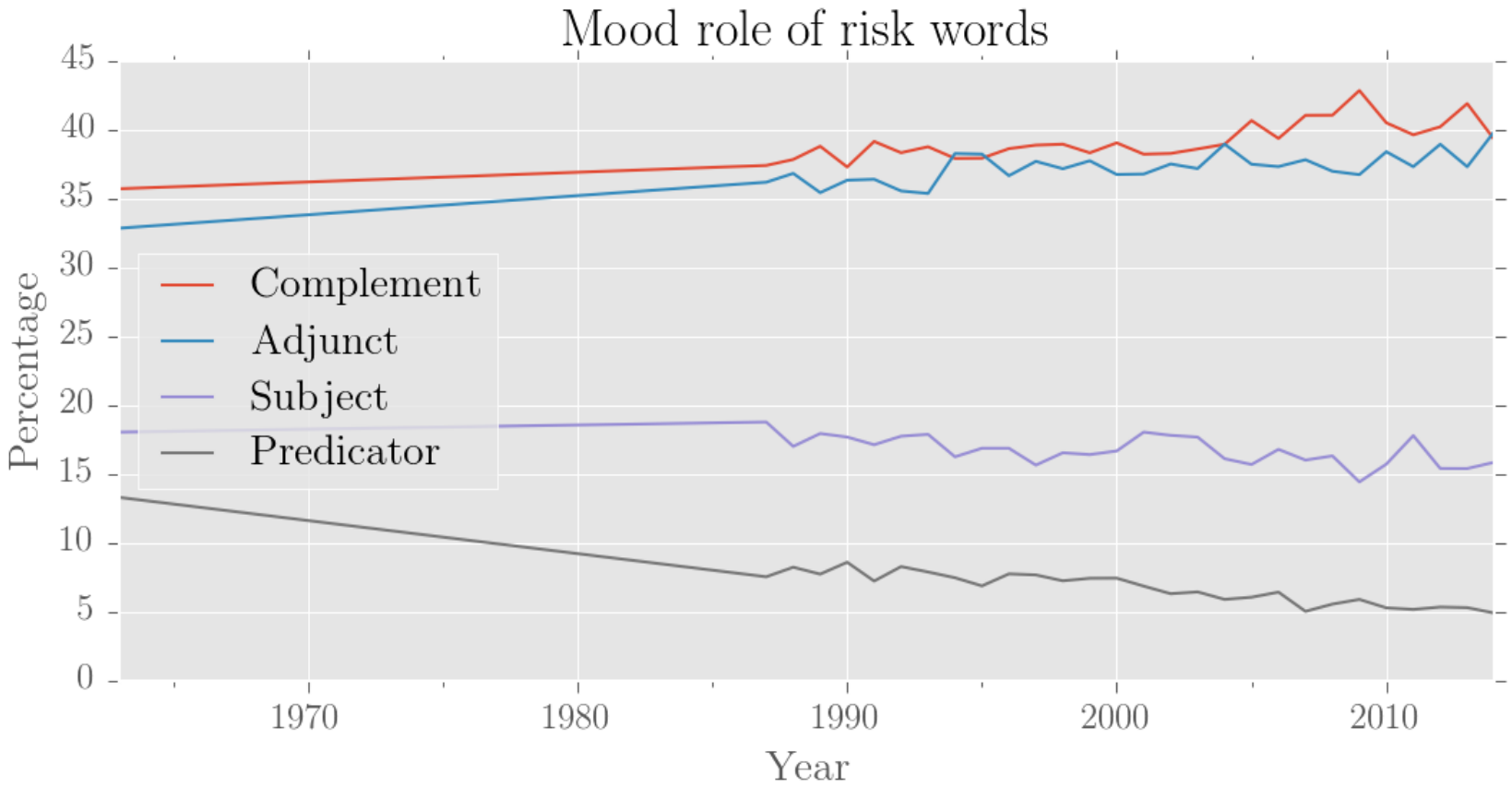


Risky decision → risk arbitrage

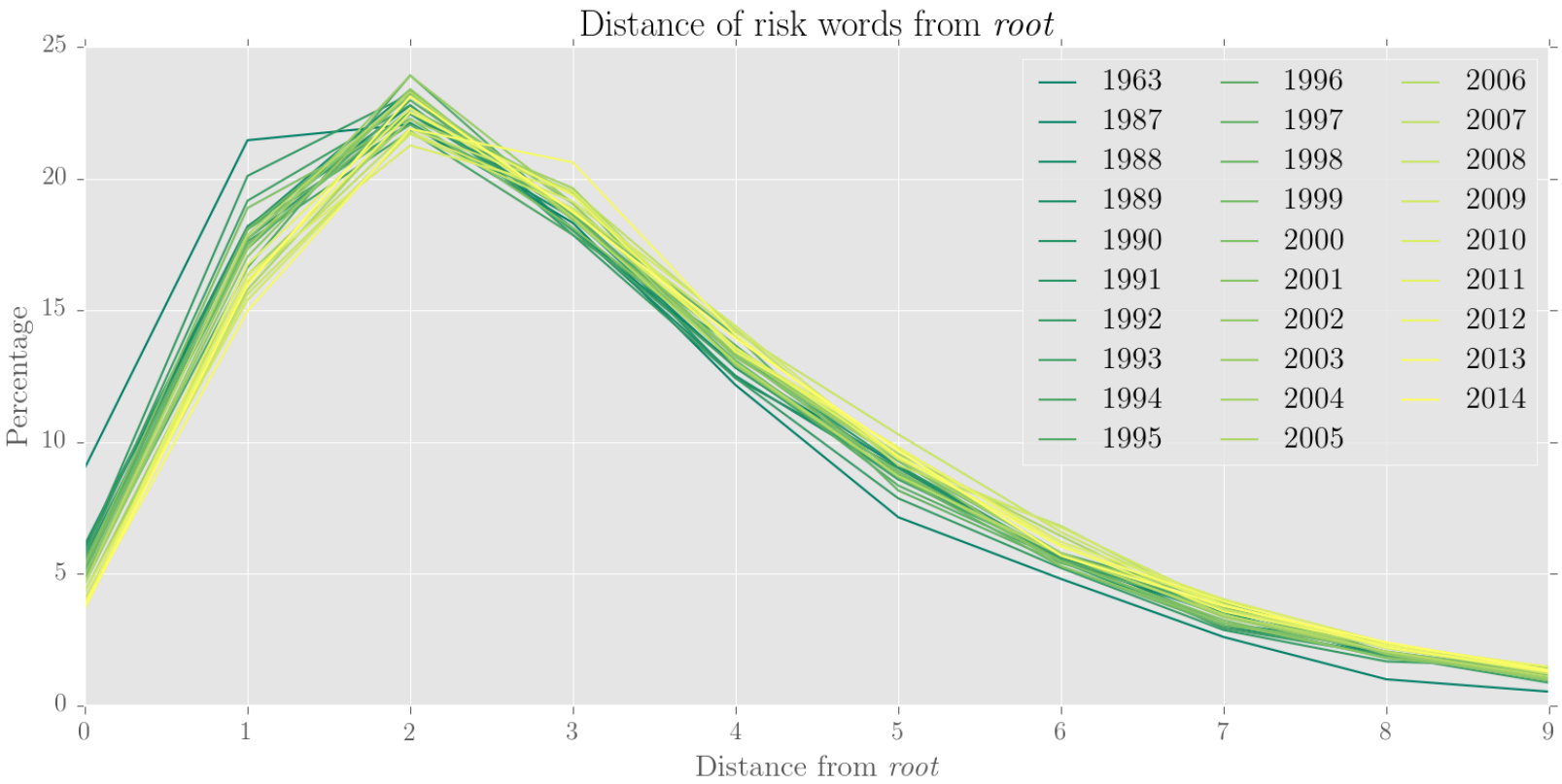
Risk and power



Mood role of risk words



Distance of risk word from *root* (predicator)



Summary of key findings

- Nominalisation and *participantification*
  - 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)

Discussion of methodology

- 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

Discussion: sociological and linguistics

THE UNIVERSITY OF MELBOURNE

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

Research agenda

- 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.

We're open source

THE UNIVERSITY OF MELBOURNE

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/vY0M4`

References I

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. (00000)

Dean, M. (1998). Risk, calculable and incalculable. *Soziale Welt*, 25–42.

Dean, M. (2010). *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*.

Foucault, M., Martin, L. H., Gutman, H., & Hutton, P. H. (1988). *Technologies of the self: A seminar with Michel Foucault*. Univ of Massachusetts Press.

References II

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

Heyvaert, L. (2003). Nominalization as grammatical metaphor: On the need for a radically systemic and metafunctional approach. *status: published*. (00012)

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://dx.doi.org/10.1080/21698252.2014.937563` doi: 10.1080/21698252.2014.937563