LSSによるアジア言語のメディア分析

量的テキスト分析を用いた報道バイアスの測定

渡辺耕平

LSSによるメディアの内容分析

ウクライナ危機におけるロシア国営メディアのバイアス (Watanabe 2017)

Latent Semantic Scalingを作った背景

- ウクライナ危機におけるロシア国営メディアを分析するため
  - 2013年9月にキエフで反政府運動が始まり，翌年3月にクリミアがロシアに併合され，東部ではウクライナ軍と民兵の戦闘が始まる
  - ロシアの通信社によって2年間で20万件のウクライナに関するニュースが発行された
- メディアの量的分析の方法が確立していない
  - バイアスを測定するためには，通時的に大量の文書を分析しなくてはいけない
  - 既存のキーワード辞書がウクライナ危機で使えるかどうかは定かではない

メディア研究における報道バイアス

- バイアスはメディア研究の中心的な問題
  - バイアスはジャーナリズムに関して一番広く議論されている問題のひとつ　(Barkho, 2013b; Donsbach & Klett, 1993; Maras, 2012)
- ニュースのバイアスの定義の不一致
  - ニュースのバイアスは客観性，中立性，平等性，公平性（objectivity, impartiality, balance, fairness）の欠如とされる
  - 実証研究では測定方法によってバイアスの定義が異なる
- バイアスの研究はアメリカのメディアの研究が中心
  - アメリカにおける選挙報道（政党，候補者，アジェンダ）が研究中心であった
  - 党派性のある国際メディアでは国内メディアと同じ方法を使えない

従来の研究のアプローチ

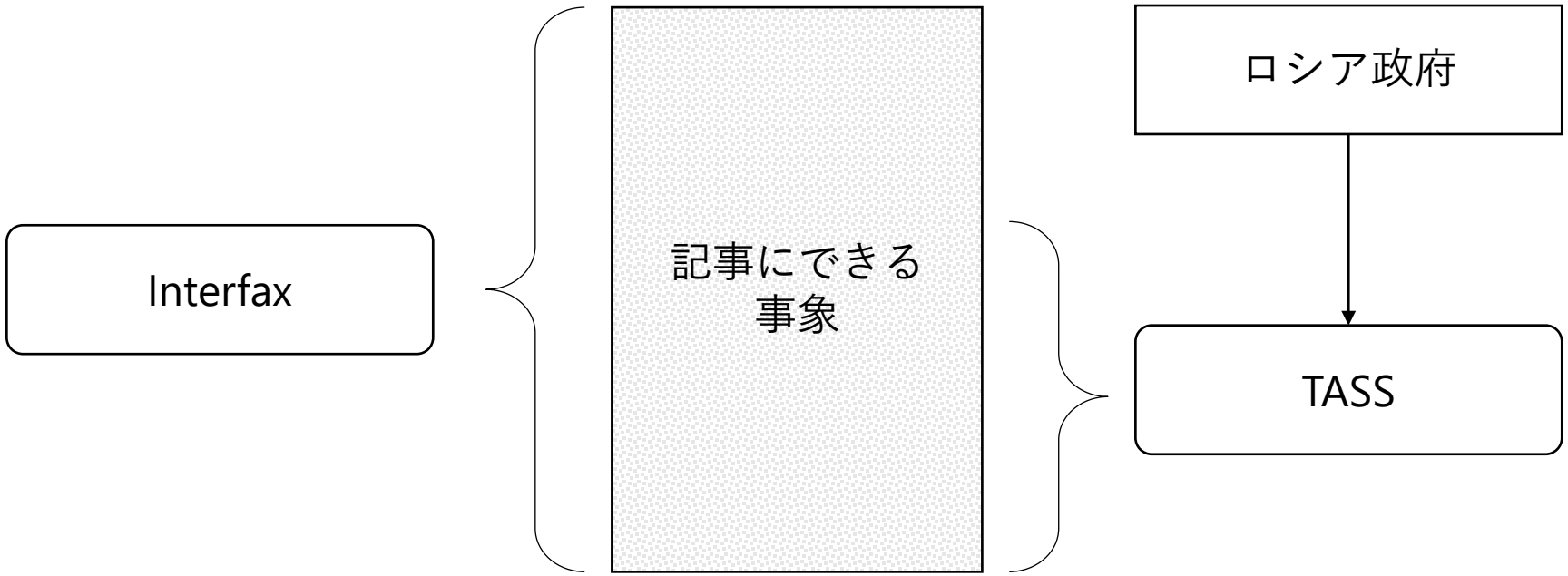
- 二等分の基準（50-50 benchmark）
  - 異なる集団や話題に関する報道のトーンや量が同じでなくては
  - アメリカの二大政党制とFCCの機会均等の原則（Fairness Doctrine）に基づく
  - メディアの多様化により機会均等の原則は1987年に廃止
- その他の基準
  - 人口の地理的な分散（Cushion, Lewis, and Groves 2009）
  - アイルランド議会における政党の割合（Brandenburg 2005）

LSSを使った研究のアプローチ

- 定義：ニュースの発行者の利益のためのメディアの利用
  - 中立性の欠如に近いが，平等性は問題にならない
  - 報道内容の客観性や正確性も問題にならない
- 測定：重要な問題に関するフレーミングの変化
  - 複数のメディアを通時的に分析し，内容の変化を相対的に分析する
    - ウクライナの民主主義と主権に関するポジティブ／ネガティブ感情
    - 中国の汚職問題に関する宣伝／番犬フレーム
  - “Consistent patterns in the framing of mediated communication that promote the influence of one side in conflicts over the use of government power” (Entman 2010)

LSSを用いた測定の論理

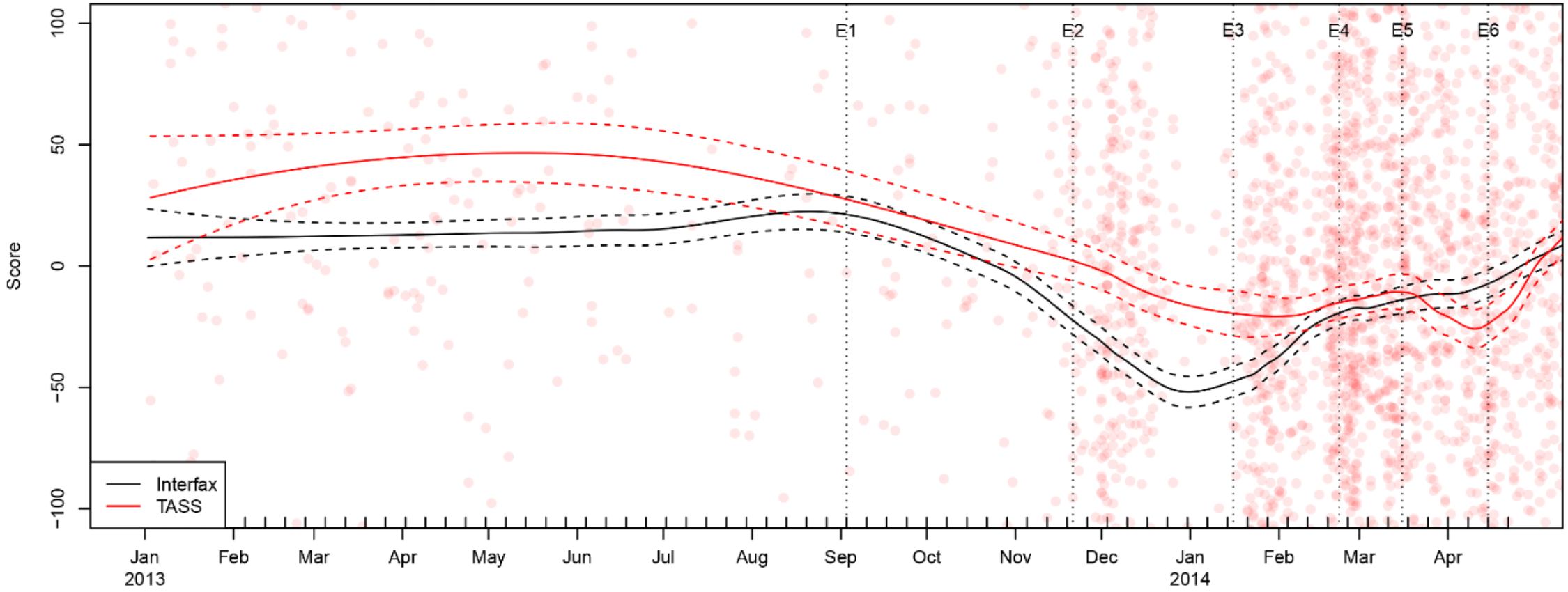
- 報道は記事にできるすべての事象から一部を選択する作業
  - 特定の集団や話題の系統的な選択が報道バイアスを生じさせる
  - 政治的な動機づけによって系統的な選択が行われる



LSSによる内容分析

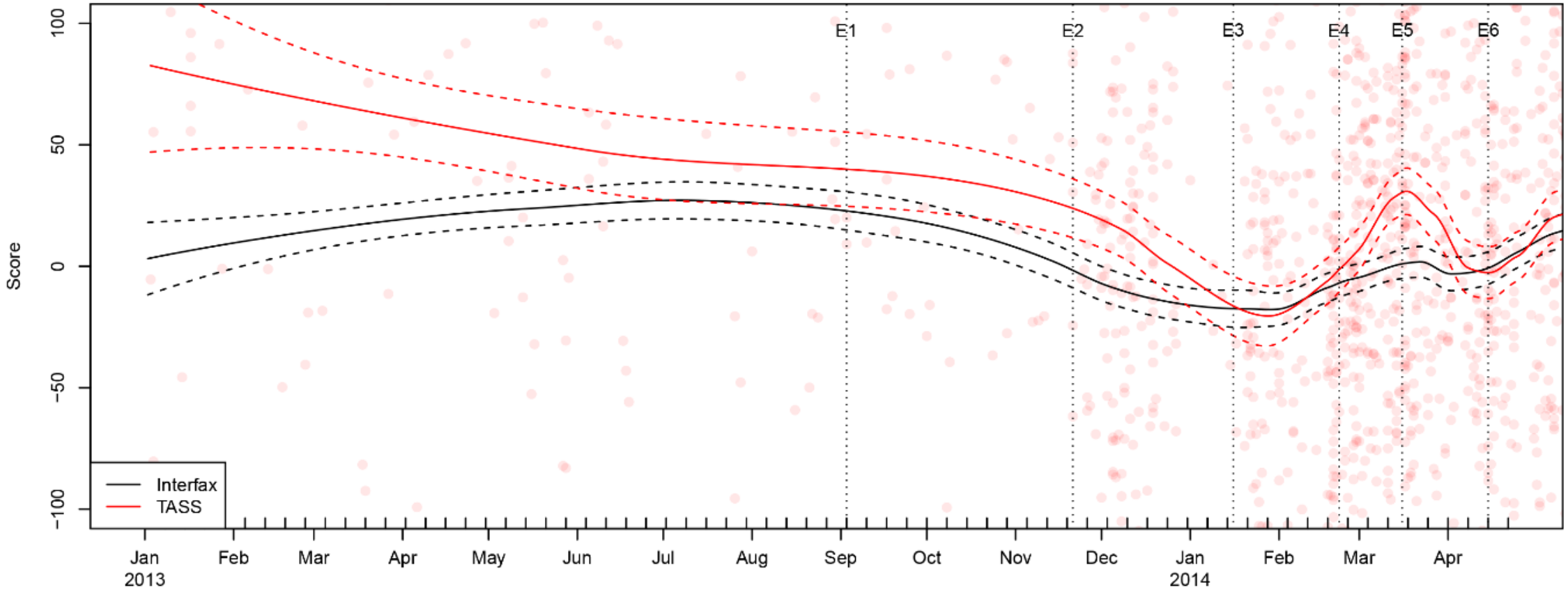
- 主題に関する記事を広く収集
  - TASSとInterfaxによるウクライナに関する英語の記事を収集
  - キーワード検索でデータベース（NexisとIntegrum）からダウンロード
- 主題に強く関係する記事をより厳密に選択
  - Newsmapを用いてウクライナでの出来事に関する記事を選択
  - キーワード辞書分析で民主主義と主権に関する記事を選択
- 選択された記事からLSSを作成
  - Quantedaの前身となるプログラムで文書の前処理を行い，文書行列を作成
  - LSSに感情種語を与え，文書行列から語の極性を推定
- LSSで記事に対して感情スコアを与える
  - ポジティブ／ネガティブ

ウクライナの民主主義のフレーミング



E1 Yanukovich demands legal reforms to MPs for EU association plan
E2 The trade agreement with the EU is abandoned by Yanukovych
E3 Protest against the pro-Russian regime in Kiev intensifies
E4 Yanukovych is removed from presidency by the parliament
E5 Crimea referendum is held and 95% support accession
E6 Military operations against separatists are launched

ウクライナの主権のフレーミング



E1 Yanukovich demands legal reforms to MPs for EU association plan
E2 The trade agreement with the EU is abandoned by Yanukovych
E3 Protest against the pro-Russian regime in Kiev intensifies
E4 Yanukovych is removed from presidency by the parliament
E5 Crimea referendum is held and 95% support accession
E6 Military operations against separatists are launched

LSSによるアジア言語の分析

日本とイスラエルの新聞報道における核脅威の比較分析 (Watanabe et al, 2022)

アジア言語の重要性

- メディア制度や政治体制の研究では国際比較が広く行われている
  - メディアや議会政治の研究では，記事の数や発言回数だけではなく，内容分析を求められる
- 英語が使われている国だけでは世界的な現象を研究できない
  - 世界的な出来事に関する研究では，同様な方法で複数の国や言語の文書を分析しなくてはならない
- 中国の台頭によってアジアの重要性が高まっている
  - 中国やその周辺の国々の文書を分析する必要がある

アジア言語の量的分析

- ソフトウェアが未発達で準備が煩雑だった
  - 前処理のために形態素解析プログラムなど外部ツールが必要だった
  - ソフトウェアの発達によって前処理から応用分析まで単一のツールでできるようになった
- アジア言語ではキーワード辞書が少ない
  - フレーミングなど特定の概念に関する分析を行えない
  - LSSは少数の種語を選択するだけで文書の極性を測定できる
    - 中国語：宣伝／番犬 (Yu, 2023)
    - 韓国語：過去／現在 (Matsuo et al., 2023)
    - アラビア語：融和／対立 (Yamao, 2019)

日本語のトークン化

トークン化で文書を単語や記号などの個別の要素に分割する

```
> txt_jp <- "政治とは社会に対して全体的な影響を及ぼし、社会で生きるひとりひとりの人の人
生にも様々な影響を及ぼす複雑な領域である。"
> quanteda::tokens(txt_jp)
tokens from 1 document.
text1 :
 [1] "政治"        "と"          "は"          "社会"        "に対して"
 [6] "全体"        "的"          "な"          "影響"        "を"
[11] "及"          "ぼ"          "し"          "、"          "社会"
[16] "で"          "生きる"      "ひとりひとり" "の"          "人"
[21] "の"          "人生"        "に"          "も"          "様々"
[26] "な"          "影響"        "を"          "及ぼす"      "複雑"
[31] "な"          "領域"        "で"          "ある"        "。"
```

中国語のトークン化

トークン化で文書を単語や記号などの個別の要素に分割する

```
> txt_cn <- "政治是各种团體进行集体决策的一个过程，也是各种团體或个人为了各自的領域所结成的特定关系，尤指對於某一政治實體的統治，例如統治一個國家，亦指對於一國內外事務之監督與管制。"
> quanteda::tokens(txt_cn)
tokens from 1 document.
text1 :
 [1] "政治"   "是"     "各种"   "团"     "體"     "进行"   "集体"   "决策"
 [9] "的"     "一个"   "过程"   "，"     "也是"   "各种"   "团"     "體"
[17] "或"     "个人"   "为了"   "各自"   "的"     "領域"   "所"     "结成"
[25] "的"     "特定"   "关系"   "，"     "尤"     "指"     "對於"   "某一"
[33] "政治"   "實體"   "的"     "統治"   "，"     "例如"   "統治"   "一個"
[41] "國家"   "，"     "亦"     "指"     "對於"   "一"     "國內外" "事務"
[49] "之"     "監督"   "與"     "管制"   "。"
```

北朝鮮が核実験

3回目 金正恩体制では初

朝日新聞

号外

北朝鮮の朝鮮中央通信は12日、「3回目の地下核実験を成功裏に実施した」と発表した。韓国気象庁は同日午前11時57分ごろに、北朝鮮北東部で「人工地震とみられる揺れ」を観測したことを明らかにした。

韓国国防省報道官は「核実験と推定している」と述べた。実施は2006年10月と09年5月に続き3回目で、金正恩（キム・ジョンウン）体制では初めてとなる。国連安全保障理事会は緊急会合を開く。

過去2回は北東部の咸鏡北道（ハムギョンブクト）・豊渓里（プンゲリ）の核実験場で実施しており、今回、揺れが観測された場所もこの周辺だという。韓国国防省報道官は会見で、北朝鮮が事前に核実験の実施を米中に通報していたとの認識を示した。

韓国の李明博（イ・ミョンバク）大統領は午後1時に国家安全保障会議を緊急に招集。朴槿恵（パク・クネ）次期大統領も情勢報告を受け、対応策を協議する。緊急核実験と確認されれば、国連安全保障理事会も緊急に招集される見通しで、朝鮮半島の緊張が高まるのは必至だ。

北朝鮮が昨年12月12日に「人工衛星の打ち上げ」として長距離弾道ミサイルの発射実験に踏み切ったのに対し、国連安保理が1月22日に制裁決議を採択、北朝鮮はこれに強く反発して同21日に、「高い水準の核実験」を実施すると予告していた。

韓国政府は衛星写真の分析などから、北朝鮮が豊渓里で核実験の準備をすでに終え、金正恩第1書記の政治判断次第でいつでも実験ができる状態になっているとみて、監視と警戒を強めていた。

韓国気象庁によると、11時57分ごろに豊渓里を含む咸鏡北道吉州郡でマグニチュード（M）5・0程度の地震を観測した。06年の核実験でMは3・6、09年はM4・4だったという。

今回の核実験は、核爆弾の小型化を目指したものとみられている。北朝鮮は、昨年12月に米本土をうかがう長距離弾道ミサイルの発射実験に成功しており、小型化への道筋がつけば米国にとっては大きな脅威になる。

また、過去2回の核実験ではプルトニウムを使ったとみられているが、北朝鮮はこの間、ウラン濃縮活動を続けてきており、今回は高濃縮ウランを使った可能性もある。

北朝鮮は昨年2月の米朝合意でウラン濃縮活動や核実験、長距離ミサイル発射の一時停止を約束し、米国は事実上の見返りとして栄養補助食品など21万トンを支援することになったが、北朝鮮は「人工衛星の打ち上げは宇宙の平和利用で、米朝合意に反しない」として4月に発射実験を強行。この時は失敗したが、昨年12月に再び発射実験をし、成功させた。（ソウル＝貝瀬秋彦）

1月28日、北朝鮮の首都平壌で開幕した朝鮮労働党第4回細胞書記大会で開幕の辞を述べる金正恩第1書記＝朝鮮通信

朝日新聞
0120-33-0893
http://www.asahi3.com
朝日新聞 DIGITAL

ידיעות אחרונות

"איראן גרעינית היא אינם ישיר וכבד עלינו"

"בתביהו, תקשיב לבכירי מערכת הביטחון"

"קשה לי לישון בלילה כשאני יודע מה קורה"

"אני מזהירה מהרפתקה מגלומנית באיראן"

שאלת איראן

דילמת התקיפה באיראן, שנידונה עד כה בחדרי חדרים, הסעירה אתמול את הכנסת ● צביקה בריזון, עמ' 2

אמנון אברמוביץ'

ראש גרעיני

בעז אופקון

ברוני החוק

השמחה והמכה

יום שלישי בלי לימודים בערי הדרום

כיפת ברזל ירדה בארים שנדד לעבר באר שבע ● מתן צחור, עמ' 4

תוכן בחלקים של ידיעות אחרונות

1 שעות/24

הום וגנ

עיר חדשה ישנה

2 חמון

קוטפים את הפירות

3 שונות

סונר חשבון

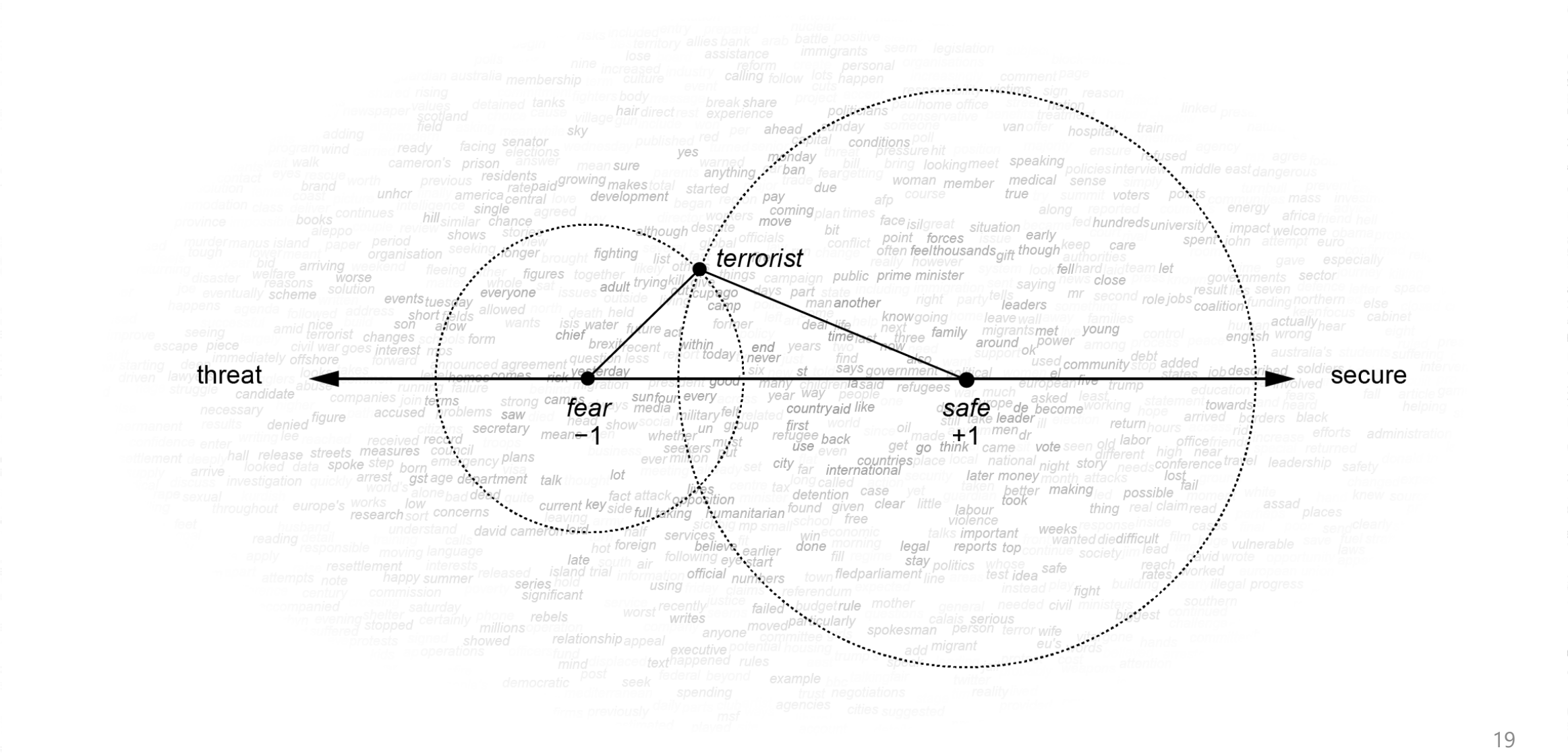
日本とイスラエルの保守系新聞の分析

- 日本とイスラエルで保守系新聞のバイアスを測定する
  - 保守政権下では保守新聞が安全保障上の脅威を強調する傾向にあるだろう（political parallelism）
  - 脅威を強調することは，政権支持を高めることに貢献するだろう（rally-around-the-flag effect）
- 安全保障に関する情報の流れは一方向であることが多い
  - 人々は新聞に依存し，新聞は政府に依存する
- LSSを使うと多言語比較分析ができることを示す
  - これまで分析が難しかった日本語とヘブライ語の分析を同じ手法で分析する

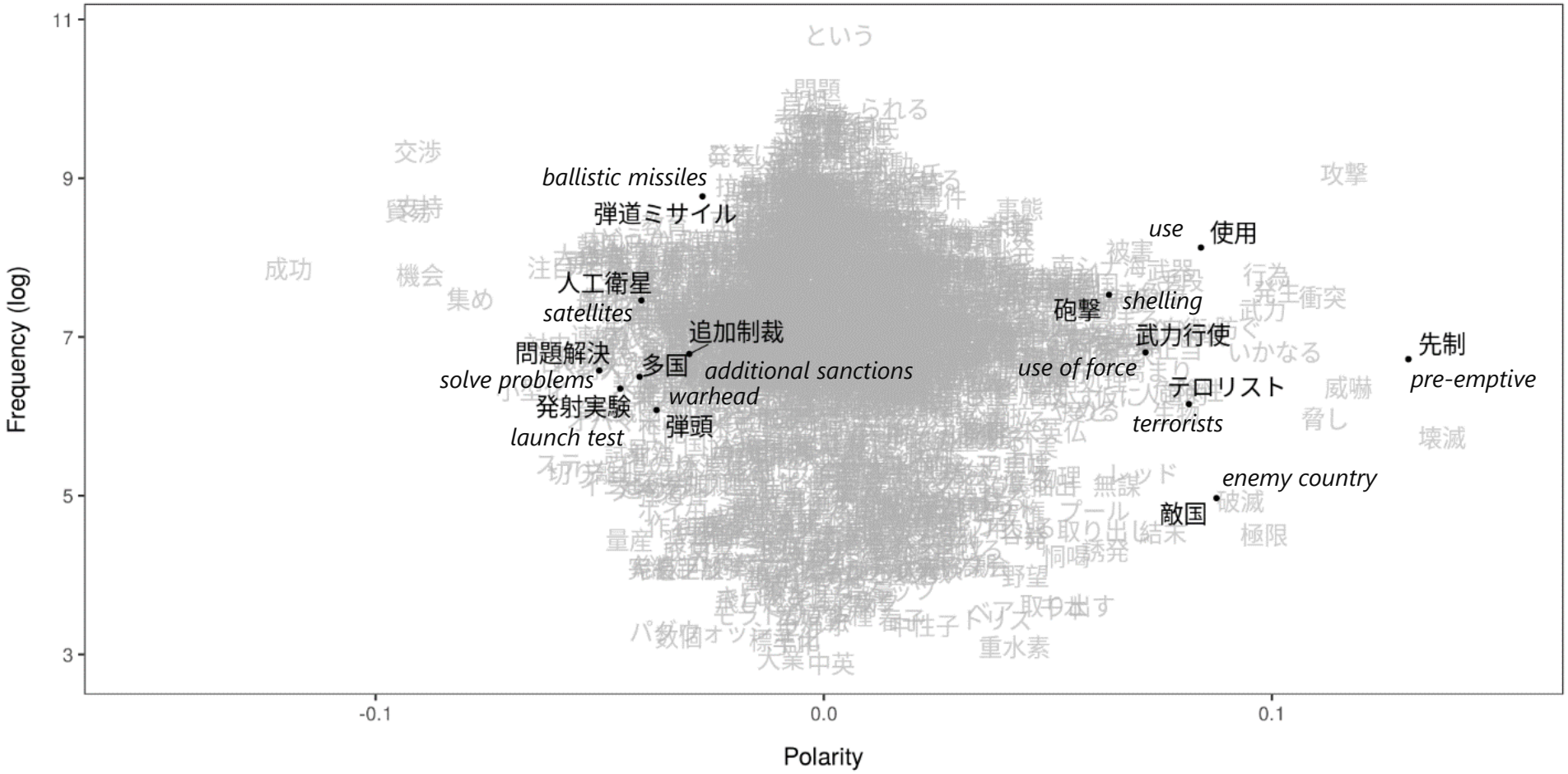
日本語とヘブライ語の種語

Polarity	Japanese	Hebrew	(English)
Threat	危険, 敵意, 壊滅, 危害, 衝突, 攻撃	סכנה, מסוכן, תוקפנות, הרס פגע, פוגע, תאונה, התנגשות, פגיעה	dangerous, hostility, complete destruction, harm/harmful, crash/collide, attack
Secure	対話, 支持, 機会, 交渉, 成功, 貿易	התקפה, שיחות, עזרה, הזדמנות, משא ומתן, הצלחה, מסחר	talk/dialog, support/assist, opportunity negotiation, success, trade

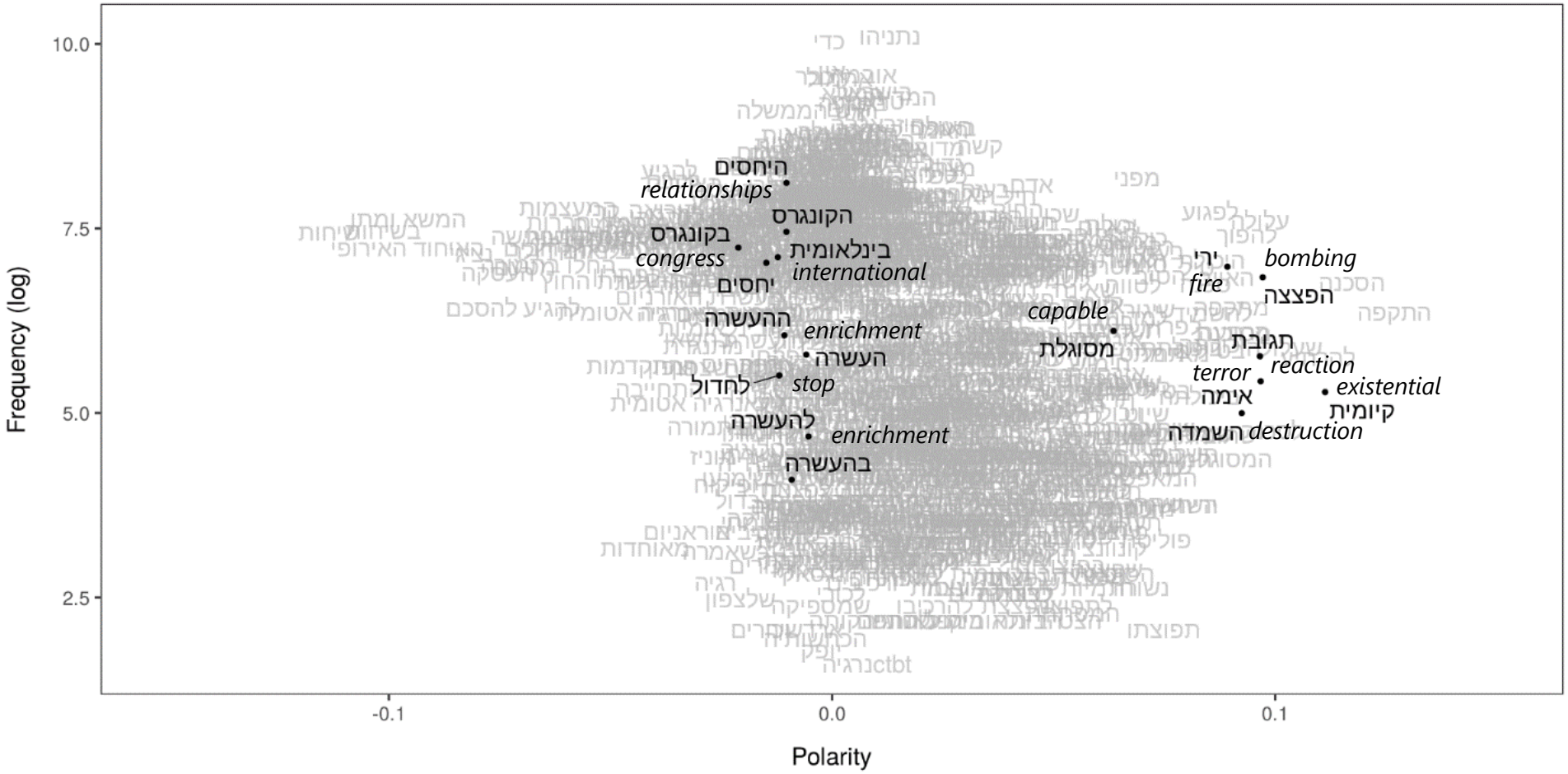
LSSによる極性判定の概念



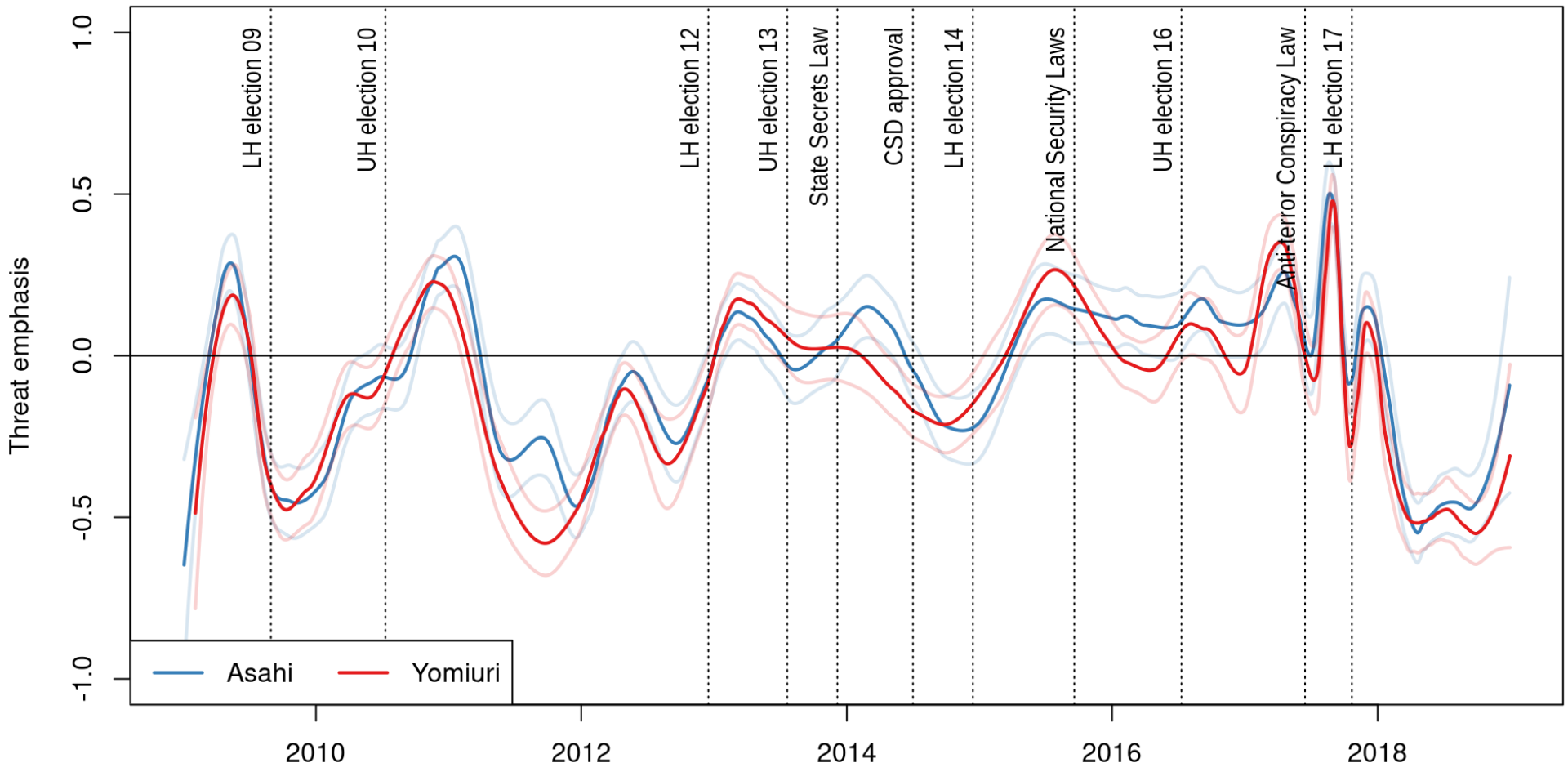
日本語の極性語


Frequency (log)
Polarity
11
9
7
5
3
-0.1
0.0
0.1
ballistic missiles
弾道ミサイル
人工衛星
satellites
追加制裁
問題解決
solve problems
多国
additional sanctions
発射実験
warhead
launch test
弾頭
use
使用
砲撃
shelling
武力行使
use of force
テロリスト
terrorists
先制
pre-emptive
enemy country
敵国
という
交渉
成功
機会
集め
攻撃
威嚇
脅し
壊滅
破滅
極限
重水素

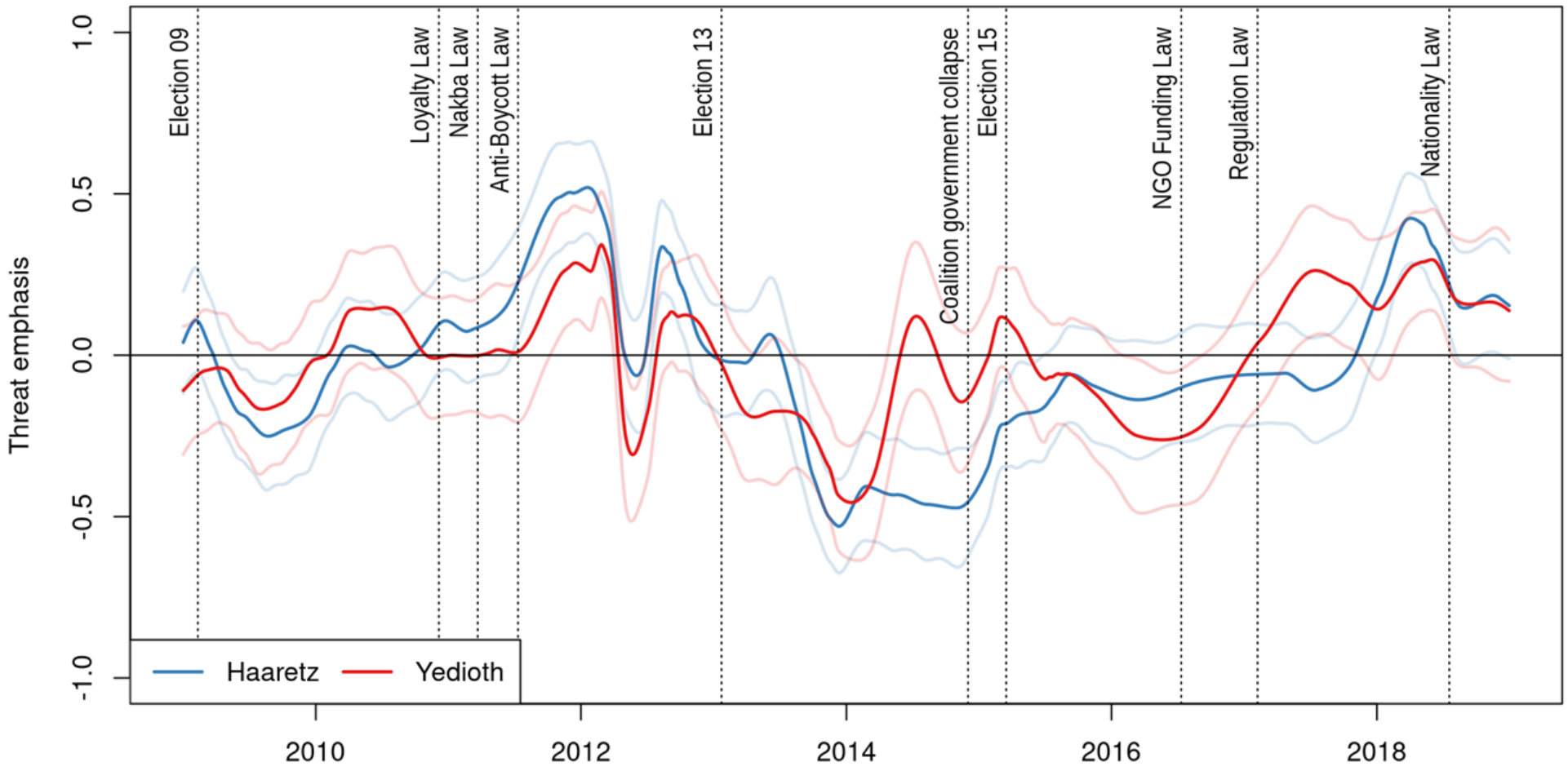
ヘブライ語の極性語


Frequency (log)
10.0
7.5
5.0
2.5
Polarity
-0.1
0.0
0.1
היחסים
relationships
הקונגרס
בקונגרס
congress
בינלאומית
international
יחסים
ההעשרה
enrichment
העשרה
לחדול
stop
להעשרה
enrichment
בהעשרה
capable
מסוגלת
ירי
fire
bombing
הפצצה
תגובת
reaction
terror
אימה
existential
קיומית
השמדה
destruction

日本の新聞


Threat emphasis
1.0
0.5
0.0
-0.5
-1.0
LH election 09
UH election 10
LH election 12
UH election 13
State Secrets Law
CSD approval
LH election 14
National Security Laws
UH election 16
Anti-terror Conspiracy Law
LH election 17
Asahi
Yomiuri
2010
2012
2014
2016
2018

イスラエルの新聞


Threat emphasis
1.0
0.5
0.0
-0.5
-1.0
Election 09
Loyalty Law
Nakba Law
Anti-Boycott Law
Election 13
Coalition government collapse
Election 15
NGO Funding Law
Regulation Law
Nationality Law
Haaretz
Yedioth
2010
2012
2014
2016
2018

LSSの課題と展望

- トピックモデルと組みわせたより精査な分析ができる
  - Sequential LDAと組み合わせると文節単位の分析ができるようになる (Watanabe & Baturo, 2023)
- 種語の選択をより簡単にする
  - 研究者が手で種語の選択を行うべきだが，その良し悪しを検証する手法を開発する必要がある
- 文書レベルでの測定エラーを小さくする
  - 教師あり学習モデルと比較すると，測定エラーがかなり大きい
- パラメーターの最適化を行えるようにする
  - LSAの次元数（200 ≤ k ≤300）を最適化する方法を開発する必要がある

参考文献

- Watanabe, K. (2017). Measuring news bias: Russia's official news agency ITAR-TASS' coverage of the Ukraine crisis. *European Journal of Communication*, *32*(3), 224–241. https://doi.org/10.1177/0267323117695735
- Watanabe, K., Segev, E., & Tago, A. (2022). Discursive diversion: Manipulation of nuclear threats by the conservative leaders in Japan and Israel: *International Communication Gazette*. https://doi.org/10.1177/17480485221097967
- Watanabe, K., & Baturo, A. (2023). Seeded Sequential LDA: A Semi-Supervised Algorithm for Topic-Specific Analysis of Sentences. *Social Science Computer Review*. https://doi.org/10.1177/08944393231178605
- Yamao, D. (2019). *ISのインパクトをはかる：イラク主要3紙の量的テキスト分析から [Measuring the Imapct of Isramic States: Quantitative Analysis of Three Iraqi newspapers]* (Relational Studies on Global Crises). Chiba University. http://www.shd.chiba-u.jp/glblcrss/online_papers/onlinepaper20190819.pdf
- Matsuo, A., Han, O., & Matsumura, N. (2023). The social media audience of diplomatic crisis. *The British Journal of Politics and International Relations*, 13691481231172346. https://doi.org/10.1177/13691481231172346