

A Workflow for Integrating Close Reading and Automated Text Annotation

Maciej Janicki, Eetu Mäkelä, Anu Koivunen, Antti Kanner, Auli Harju, Julius Hokkanen, Olli Seuri

Flows of Power

Motivation

The Flows of Power project seeks to:

1. understand media as a site and agent in politics, and how these have changed in the last 20 years
2. develop novel research of *interest to media studies*
3. in pursuit of the above, *develop better big data research workflows*

Particularly, we've sought means to efficiently:

1. ensure our computational indicators capture what is relevant
2. combine quantitative and qualitative analyses to make sense of the complex phenomena, and
3. ensure everyone in the project shares a common understanding

Current work

Our workflow consists of three different strategies:

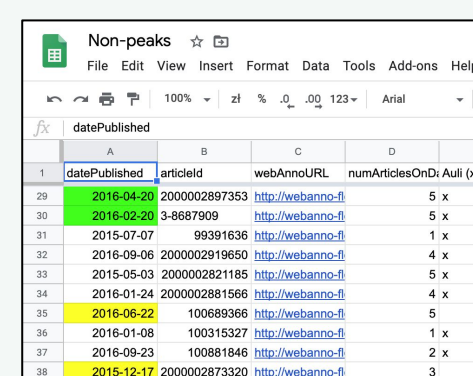
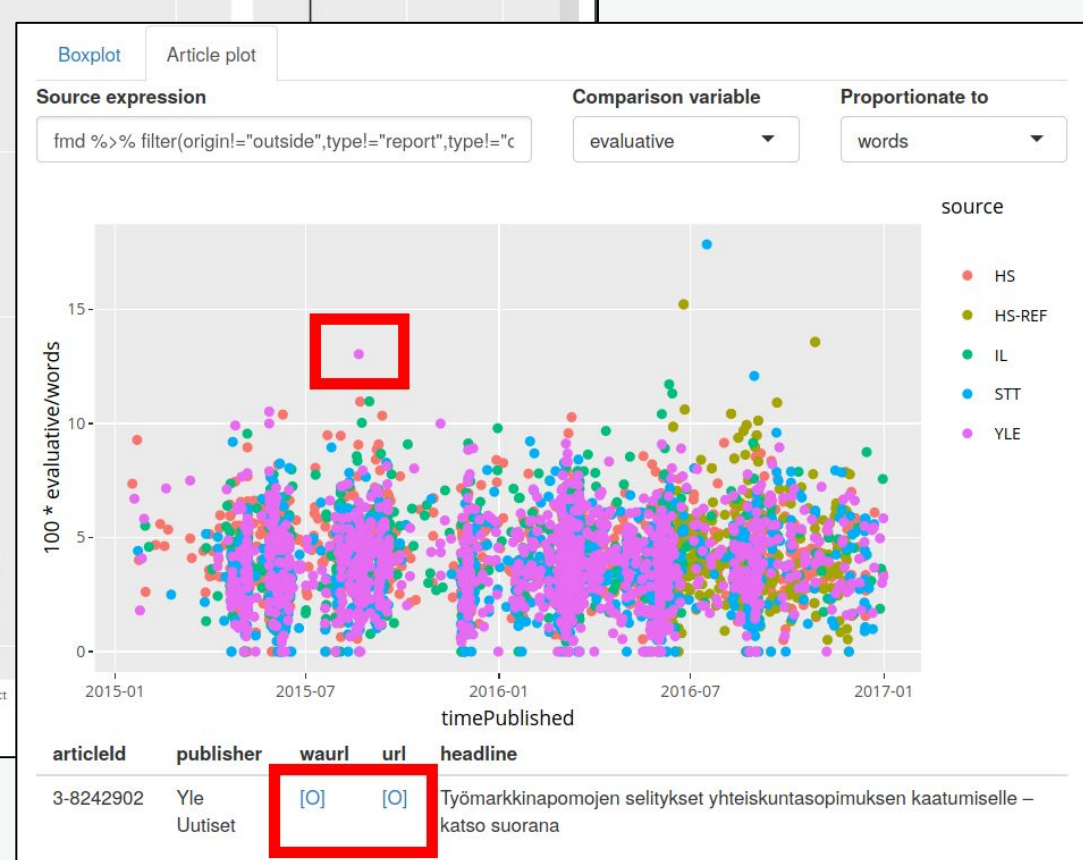
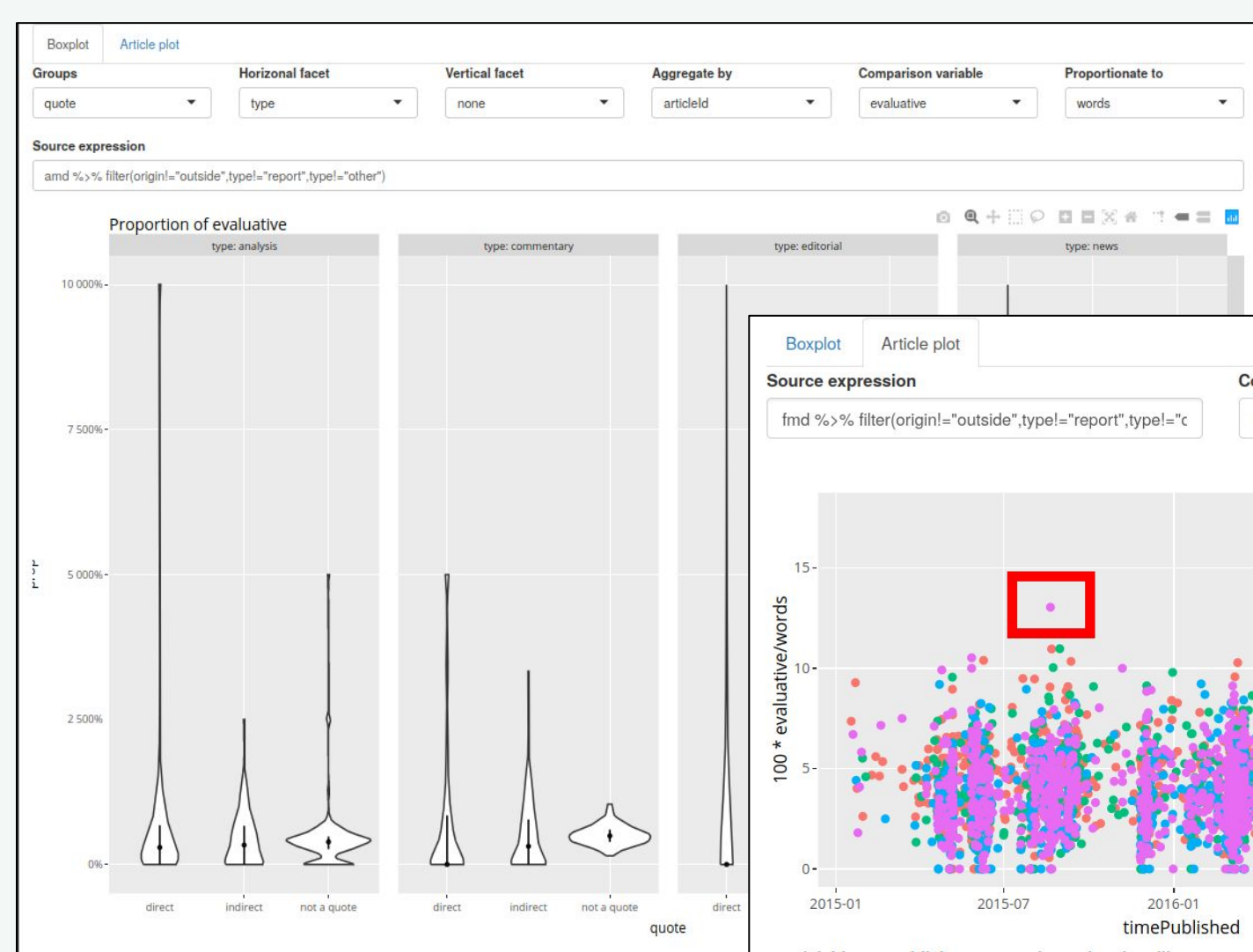
1. close reading of news articles by domain experts
2. automated text annotation
3. statistical analysis / distant reading

The efficient **flow of information between those three strategies** is a major challenge for the project. This includes questions like:

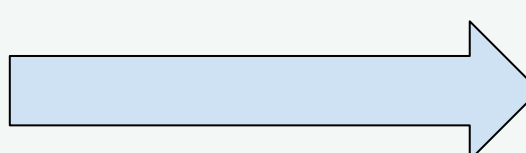
- visualization of automatic annotations during close reading
- exporting manually done annotations to a machine-readable format
- selecting articles for close reading based on statistical patterns
- developing indicators for automated annotation iteratively

Key design decisions

- Use **CSV** as data interchange format
- Avoid monolithic toolboxes. Instead, link together **general-purpose, mature** tools:
 - **R/Shiny** for statistical analysis and plotting
 - **WebAnno** for annotation and close reading
 - **Google Sheets** for viewing and manipulating tables
 - **Prolog** for rule-based annotation (esp. based on dependency trees)
 - **Turku Neural Parser** as NLP pipeline
- Prefer rule-based and wordlist-based annotation over machine learning (reason: interpretability!)



close reading



Ideas from close reading

7	Lue tästä hänen toukokuussa Helsingin Sanomissa julkaistu haastattelunsa.	TimexTimeDat	EnamexOrgTvr
8	Elorannan mukaan työn muuttuminen tarkoittaa sitä, että ay-liikkeen pitää uudistua.	EnamexPrsHum	direct modal_verb threat
9	Hänen mielestään pelisääntöjä ja sosiaaliturvaa pitää muuttaa patkaiden hallitsemalle työelämälle sopivaksi.	(Metaphor)	modal_verb threat direct (Metaphor)
10	"Tarvitsemme nimenomaan molemmat.	direct	
11	Pelkästään sosiaaliturvaa muuttamalla asioita ei saada korjattua - eikä veronmaksajilla olisi siihen varaakaan", hän sanoi valintansa jälkeen Tampereella tiistaina.	EnamexLocPpl	
12	"Uuden työn tekijät eivät näy ay-liikkeen tutkassa.	direct	
13	Tutkasignaalin ulottumattomissa kehittyi kuitenkin ilmiötä, jotka vaikuttavat myös nykyisiin jäseniimme.	direct	
14	Ay-liikkeen pitääkin hakea uudenlaista työtä tekevät mukaan edunvalvontaan ja parantaa heidän asemaansa."	EnamexOrgCnp	modal_verb threat direct
15	"Pätkä-, silppu- ja osa-aikatyön määrä sekä nollasopimusten jatkuva kasvu ovat ilmiötä, joihin on puuttuttava.	(Met)	

aggregate visualization

visualization of individual annotations

Ideas from close reading

Ideas from aggregate analysis

```
% AUTHOR sanoo, että PROPOSITION
quote_cue(AuthorHead, Cue, PropHead, false) :-
  lemma(Cue, CueLemma),    speech_act_verb(CueLemma),
  head(AuthorHead, Cue),    deprel(AuthorHead, nsubj),
  head(PropHead, Cue),      deprel(PropHead, ccomp).

% PROPOSITON, sanoo AUTHOR
quote_cue(AuthorHead, Cue, PropHead, true) :-
  lemma(Cue, CueLemma),    speech_act_verb(CueLemma),
  head(AuthorHead, Cue),    deprel(AuthorHead, nsubj),
  head(Cue, PropHead),      deprel(Cue, parataxis).

% AUTHORn mukaan PROPOSITION
quote_cue(AuthorHead, Cue, PropHead, false) :-
  lemma(Cue, "mukaan"),    deprel(Cue, case),
  head(AuthorHead, PropHead), deprel(AuthorHead, obl).

% AUTHORn mielestä PROPOSITION
quote_cue(AuthorHead, Cue, PropHead, false) :-
  lemma(Cue, "mieli"),      feats(Cue, case(ela)),
  head(AuthorHead, Cue),    deprel(AuthorHead, nmod:poss),
  head(Cue, PropHead),      deprel(Cue, obl).
```

manual annotations for evaluation

```
1,1,Tämä,tämä,PRON
1,2,on,olla,VERB
1,3,vain,vain,ADV
1,4,pieni,pieni,ADJ
1,5,esimerkki,esimerkki,NOUN
1,6,...,PUNCT
-----
1,3,5,some_annotation
```

```
#Text=Tämä on vain pieni esimerkki.
1-1 0-4 Tämä tämä PRON -
1-2 5-7 on olla VERB -
1-3 8-12 vain vain ADV *[1]
1-4 13-18 pieni pieni ADJ *[1]
1-5 19-28 esimerkki esimerkki NOUN *[1]
1-6 28-29 . PUNCT -
```

