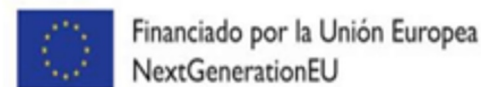




UNIVERSIDAD
DE MÁLAGA



UNIVERSITÀ
DEGLI STUDI
DI BERGAMO



Towards an annotation schema of financial discourse based on functional discourse analysis

14^o Congreso Internacional de Lingüística de Corpus
Universidad de Oviedo, Mayo 2023

Javier Fernández Cruz

Universidad de Málaga - Universidad de Bergamo

Irina Muñoz Toala

Universidad de Málaga

INTRODUCTION

Initial motivation: Lingmotif



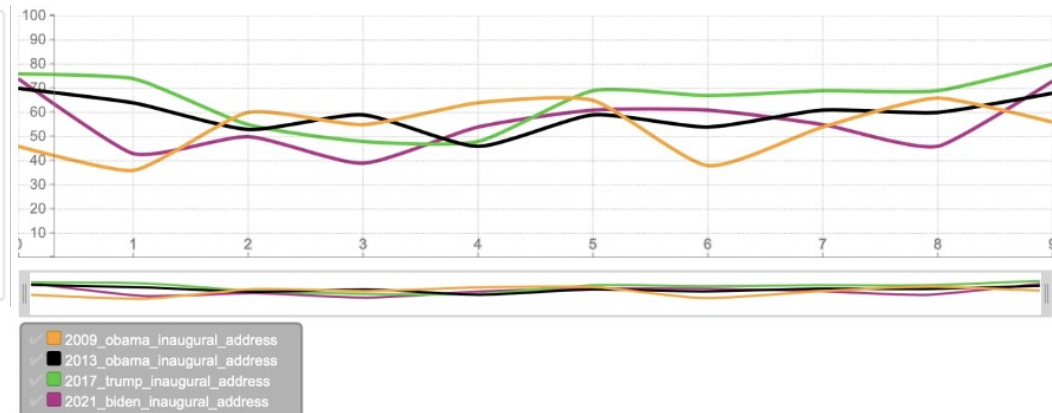
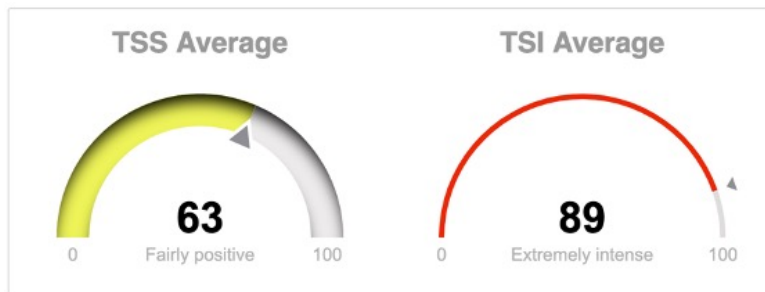
UNIVERSIDAD
DE MÁLAGA

| **uma.es**

Lingmotif is a multilingual, cross-platform, lexicon-based sentiment analysis (SA) tool.

Lingmotif 2 (Moreno-Ortiz, 2021) developed by the Tecnolengua team @ University of Málaga

It's publicly available @ [Lingmotif.com](https://lingmotif.com) (try it!)



- **Outperforms** other SA lexicons available (Moreno-Ortiz and Pérez-Hernández, 2018).
- **Highly customizable:** Users can adapt the lexicon to obtain slightly better results: e.g., SentiEcon (2020) for the economic domain.
Even better results

THE PROBLEM

At the lexical level, we obtain only marginal improvements in terms of precision.

Consumer purchasing power remained high despite a temporary **reduction in productive capacity**.

INTRODUCTION

Lingmotif at the crossroads



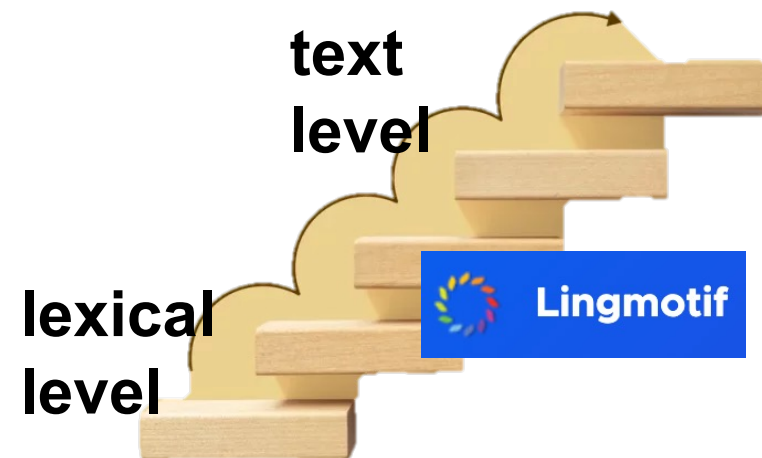
UNIVERSIDAD
DE MÁLAGA

| uma.es

THE CHALLENGE

- Better SA results
- Data-driven description evaluative discourse

Consumer purchasing power
remained high
despite a temporary reduction in productive capacity.



INTRODUCTION

A new project



UNIVERSIDAD
DE MÁLAGA

| uma.es

DisparSA

Linguistic-based sentiment analysis with rhetorical-discursive parsing



- *DisparSA* is a publicly funded project that aims to develop a SA solution at the textual level.
- Univ. Málaga, Politécnica de Valencia, Bergamo.

What we need

- ❖ Textual level framework: no previous work.
- ❖ Creation of multi-level large annotated corpora.
- ❖ As a starting point we'll work on **tourism reviews**.
- ❖ Other domains: my research targets to work with **Economy news**

OBJECTIVES (for this paper)



UNIVERSIDAD DE MÁLAGA

uma.es

To explore the specifics of the textual structure of economic opinion news texts

"All the News That's Fit to Print"

The New York Times

TODAY, CLEAR, SOME SUNSHINE, HIGH 65, LOW 55. TEMPERATURES, CLOUDY, WIND, LOW 71. TEMPERATURES, CLOUDY, SHOWERS AND THUNDERSTORMS, HEAVY AFTERNOON, HIGH 84. WEATHER MAP, PAGE 21.

VOL. CLXXI ... No. 59,487 © 2012 The New York Times Company NEW YORK, SUNDAY, JULY 17, 2012 \$6.00

Abortion Bans Leaving Young At Higher Risk

Case of 10-Year-Old Ohio Shows Barrier

By DANA GOLDSTEIN and ANA KASAN

She was just 10 years old, young that many people were horrified when they heard it, and a fetus refused to believe it. But it's not the child rape victim Ohio who had to cross state to get an abortion, and the ugly political fight that followed, have highlighted two uncomfortable facts: Such pregnancies are not as rare as people think, and new abortion bans are likely to have a pronounced impact on the young pregnant girls.

New bans in nearly a dozen states do not make exceptions for rape or incest, leaving young adolescents — already among the most restricted in their abort options — with less access to procedures. Even in states with exemptions for rape and incest, requirements involving police reports and parental consent can be prohibitive for children and teenagers.

"The situation out of Ohio is no way unique," said Katie A. High, an OB-GYN in Indiana, a board member of the group Plucians for Reproductive Health which favors abortion rights. "This is a situation that ev abortion provider has seen

RISK OF RECESSION GROWING, DRIVEN BY TIRELESS VIRUS

Home of ECONOMY

Forks, N.D., is home to several agricultural facilities, including the only state-owned flour mill in the United States.

FINANCIAL TIMES

TUESDAY 14 JANUARY 2012 NEWS PROVIDER OF THE YEAR UK £200 (Charter House £220, Republic of Ireland £220)

Victory in Taipei: Hong Kongers are taking a leaf out of Taiwan's book — NOTEBOOK, PAGE 10

The unlucky country: Australia's climate stance is an act of self-harm — GREEN RACHMAN, PAGE 11

Salvaging Salvini: Can Italy's rightwing populist make a comeback? — BIG READ, PAGE 9

Contracting economy adds to pressure for rate cut

November's GDP shows 0.3% decline
Sterling slumps to below \$1.50 mark

Under the volcano Philippines fears powerful eruption as Taal spews lava

Labour hospitals whittled down to five
Boris Johnson says he will not be a Conservative MP

EU trade chief seeks to ease US tensions

A tale of three houses

And what their prices say about Britain

Zoe Williams How can Holly and Phil rescue their reputations?

The Guardian

News papers of the year

Starker: Labour is party of the centre

Labour's mission indocrinate

New Corn Mill Brings Jobs, and Fears of China

Grand Forks, population 59,000, had revealed two hour milestone

Sterling crisis deepens as Truss's strategy unravels

Point falls to record low against dollar after Bank of England rules out immediate interest rate cut to halt slide

Concern over human rights after Italy vote

THEORETICAL FRAMEWORK

Functional Discourse Units (FDUs)



UNIVERSIDAD
DE MÁLAGA

| uma.es

We based our annotation on Egbert et al.'s (2021) **Functional Discourse Units (FDUs)**:

- any text span that is characterized by its discourse function. When we speak:

we joke, we contextualize, we advice, we share feelings...

- contiguous units in conversational language that are characterized by coherent communicative purposes

FDUs must be:

- Recognizably self-contained
- Sequentially bounded with detectable beginnings and ends
- Thematically and functionally coherent.

RESEARCH DESIGN

(What we did)



UNIVERSIDAD
DE MÁLAGA

| uma.es

1. Compile a corpus of Economy opinion columns from English language newspapers
2. Generate a dataset to serve as a basis for improving sentiment analysis at the textual level
3. Generate comprehensive set of Functional Discourse Units for this register.
4. Manual annotation of the corpus
5. Provide descriptive statistics

RESEARCH DESIGN

The Eco-DisparSA Corpus



UNIVERSIDAD
DE MÁLAGA

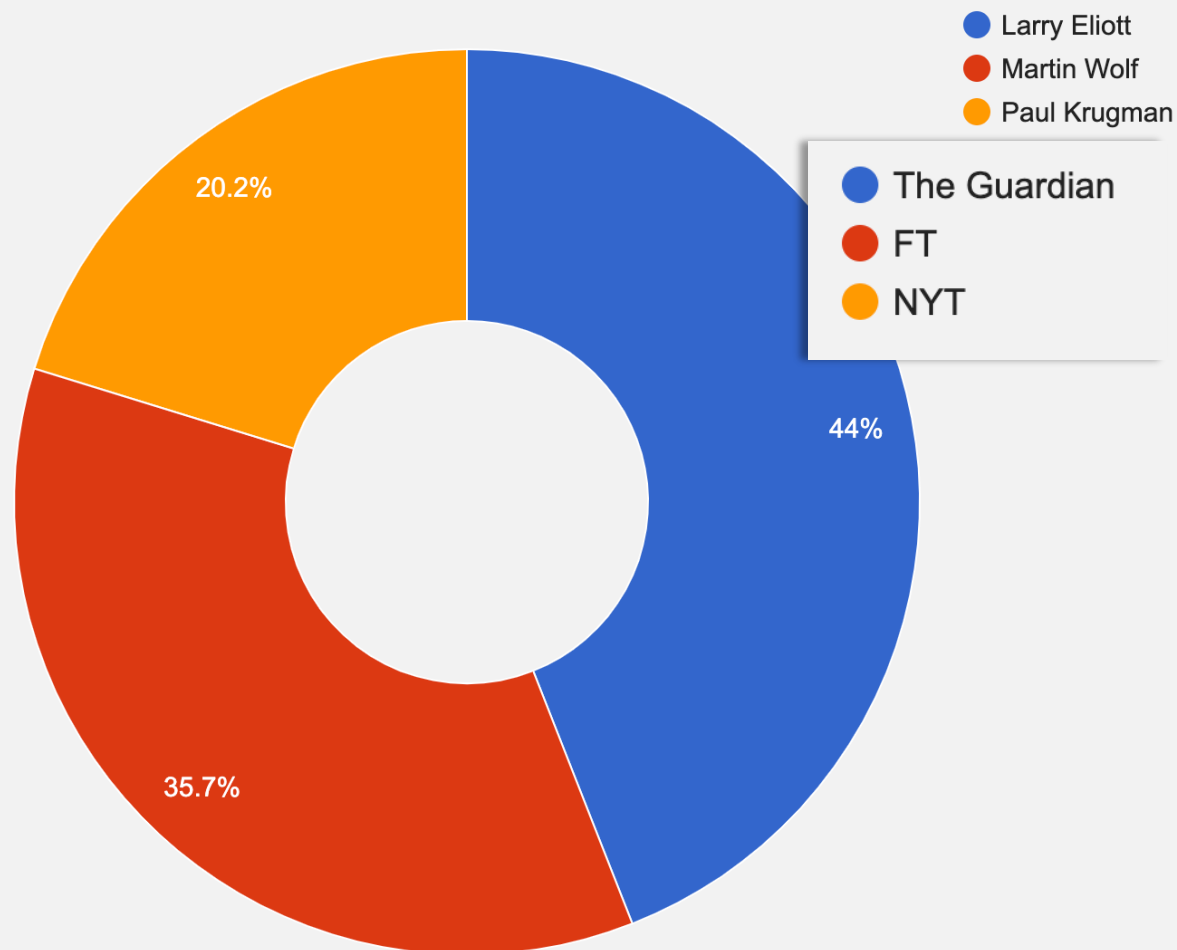
| uma.es

Tokens	80.305
Words	70.003
Sentences	3532
Documents	85
Dates	01/10/2022 20/12/2022

RESEARCH DESIGN

The Eco-DisparSA Corpus

xml - author





UNIVERSIDAD DE MÁLAGA

uma.es

RESEARCH DESIGN

The Eco-DisparSA Corpus



RESEARCH DESIGN

Tools



UNIVERSIDAD
DE MÁLAGA

| uma.es

Prodigy is a corpus annotation tool based in Python
Highly customizable
It has a friendly user interface and works in your browser



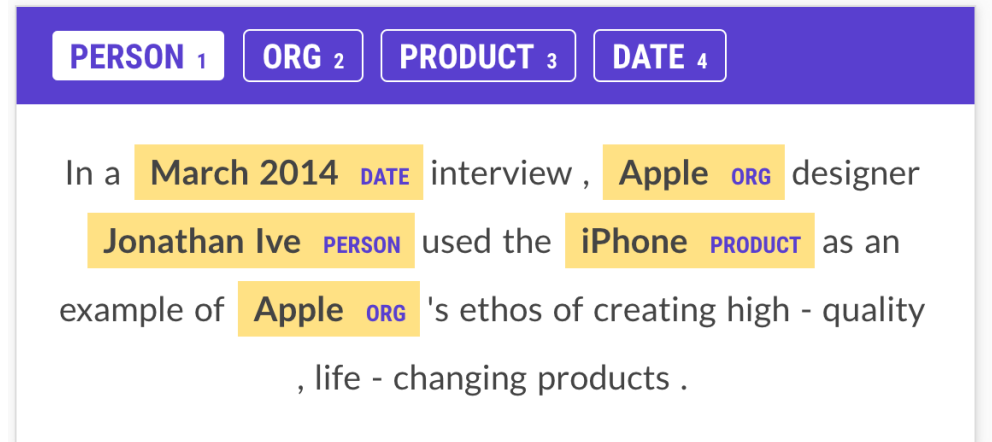
Applications

Named Entity Recognition, Span Categorization, Text Classification, Dependencies & Relations, Evaluation

```
⌘$ pip install ./prodigy.whl
100%

⌘$ prodigy ner.manual reviews_ner en_core_web_sm
./data.jsonl --label PRODUCT,PERSON,ORG

🌟 Starting the web server on port 8080...
Open the app in your browser and start annotating!
```



RESEARCH DESIGN

Our annotation framework



UNIVERSIDAD
DE MÁLAGA

uma.es

LEX:POS 1 LEX:NEG 2 FDU:CONTEXT 3 FDU:DESCRIPTION 4
FDU:ADVICE 5 FDU:EVAL_NEG 6 FDU:EVAL_POS 7
FDU:PRED_POS 8 FDU:PRED_NEG 9 ENT:OP HOLDER 10
ENT:OP HOLDER_OTHER 11 ENT:EVAL_ENTITY 12
ENT:OTHER_EVAL_ENTITY 13 ENT:READER 14

The optimists were right and can be so again ↵

FDU:EVAL_POS

LEX:POS

We cannot wait for another period of catastrophe before we

FDU:ADVICE

LEX:NEG

attempt renewal ↵

Is our world getting better and likely to continue to do so or is it

FDU:CONTEXT

sitting on the edge of catastrophe? People who think about these

LEX:NEG

This is how our
annotation schema
looked on Prodigy

RESEARCH DESIGN

Our annotation framework



UNIVERSIDAD
DE MÁLAGA

| uma.es

Use of Prodigy to explore and annotate text segments

- (1) Polarity
- (2) Functional Discourse Units
- (3) Entity
- (4) Opinion Holder

RESEARCH DESIGN

Our annotation framework



UNIVERSIDAD
DE MÁLAGA

uma.es

- (1) Polarity
- (2) Functional discourse units
- (3) Entity
- (4) Opinion Holder

POSITIVE / NEGATIVE

When good economic news may not be good news ←

FDU:DESCRIPTION

LEX:POS

LEX:POS

But while I like the idea of a “moron” premium, there may also be a

FDU:EVAL_POS

ENT:OP HOLDER

LEX:NEG

FDU:EVAL_NEG

more concrete concern. I've been in correspondence with other City

FDU:CONTEXT

LEX:NEG

RESEARCH DESIGN

Our annotation framework



UNIVERSIDAD
DE MÁLAGA

| uma.es

(1) Polarity

(2) Functional Discourse Units

(3) Entity

(4) Opinion Holder

CONTEXT

→ Past events or recaps

DESCRIPTION / ANALYSIS

→ Current state of affairs

ADVICE / CALL TO ACTION

→ Tips by the writer

EVALUATION (POS/NEG)

→ Explicit stance on affairs

PREDICTION (POS/NEG)

→ Forecast of economy

RESEARCH DESIGN

Our annotation framework



UNIVERSIDAD
DE MÁLAGA

uma.es

(1) Polarity

(2) Functional discourse units

(3) Entity

(4) Opinion Holder

CONTEXT

→ Past events or recaps

Before I get there, a couple of notes on what the economics portion

FDU:CONTEXT

ENT:OP HOLDER

of the commitment actually says. ↵

LEX:NEG

For a long time, the G.O.P. seemed to fit the portrait famously drawn

FDU:CONTEXT

ENT:EVAL_ENTITY

by Thomas Frank in his book "What's the Matter With Kansas?"

RESEARCH DESIGN

Our annotation framework



UNIVERSIDAD
DE MÁLAGA

| uma.es

- (1) Polarity
- (2) Functional discourse units
- (3) Entity
- (4) Opinion Holder

CONTEXT

→ Past events or recaps

DESCRIPTION / ANALYSIS

→ Current state of affairs

ADVICE / CALL TO ACTION

→ Tips by the writer

EVALUATION (POS/NEG)

→ Explicit stance on affairs

PREDICTION (POS/NEG)

→ Forecast of economy

of fiscal space. In particular, a country like the United States, or for

FDU:DESCRIPTION

that matter Britain, can normally run quite big budget deficits

without creating a run on its currency. This is because investors

FDU:DESCRIPTION

RESEARCH DESIGN

Our annotation framework



UNIVERSIDAD
DE MÁLAGA

| uma.es

(1) Polarity

(2) Functional discourse units

(3) Entity

(4) Opinion Holder

CONTEXT

→ Past events or recaps

DESCRIPTION / ANALYSIS

→ Current state of affairs

ADVICE / CALL TO ACTION

→ Tips by the writer

EVALUATION (POS/NEG)

→ Explicit stance on affairs

PREDICTION (POS/NEG)

→ Forecast of economy

Given the uncertainty, there is now a good case for adopting a wait

FDU:ADVICE

LEX:POS

and see position.

RESEARCH DESIGN

Our annotation framework



UNIVERSIDAD
DE MÁLAGA

uma.es

(1) Polarity

(2) Functional discourse units

CONTEXT	→ Past events or recaps
DESCRIPTION / ANALYSIS	→ Current state of affairs
ADVICE / CALL TO ACTION	→ Tips by the writer
EVALUATION (POS/NEG)	→ Explicit stance on affairs
PREDIC	

The initial response to the inevitable U-turn was moderately

FDU:EVAL_POS

encouraging for the government. Sterling rose by about a cent

LEX:POS

FDU:DESCRIPTION

Throughout the seemingly endless Tory leadership campaign, Truss

FDU:EVAL_POS

LEX:NEG

has been radiating positivity, but the time for boosterism is over.

LEX:POS

FDU:EVAL_NEG

LEX:POS

RESEARCH DESIGN

Our annotation framework



UNIVERSIDAD
DE MÁLAGA

| uma.es

(1) Polarity

(2) Functional discourse units

CONTEXT	→ Past events or recaps
DESCRIPTION / ANALYSIS	→ Current state of affairs
ADVICE / CALL TO ACTION	→ Tips by the writer
PREDICTION (POS/NEG)	→ Forecast of economy

falling exports and rising imports will eventually be a major economic drag.

FDU:PRED_NEG

Of the other G7 countries, Japan is expected to post the strongest

FDU:PRED_POS

growth (1.6%) followed by Canada (1.5%) and the US (1%).

RESEARCH DESIGN

Our annotation framework



UNIVERSIDAD
DE MÁLAGA

uma.es

- (1) Polarity
- (2) Functional discourse units
- (3) Entity
- (4) Opinion Holder

EVALUATED ENTITY
OTHER ENTITY

encouraging for the government. Sterling rose by about a cent

LEX:POS

FDU:EVAL_POS

ENT:EVAL_ENTITY

against the dollar and bond yields – the interest rates paid to the

FDU:EVAL_POS

ENT:OTHER_EVAL_ENTITY

investors who finance new UK borrowing – fell back a little. Signs of

RESEARCH DESIGN

Our annotation framework



UNIVERSIDAD
DE MÁLAGA

uma.es

(1) Polarity

(2) Functional discourse units
(3) Entity

(4) Opinion Holder

OPINION HOLDER
OTHER OPINION HOLDER

Some such energy package was necessary, for reasons I laid out two

FDU:EVAL_POS

x ENT:OP HOLDER

Truss and her chancellor, Kwasi Kwarteng, say Treasury “orthodoxy”

FDU:EVAL_NEG

ENT:OP HOLDER_OTHER

ENT:OP HOLDER_OTHER

A script generates a report from the annotated corpus:

- for every text
- aggregated info for the whole dataset

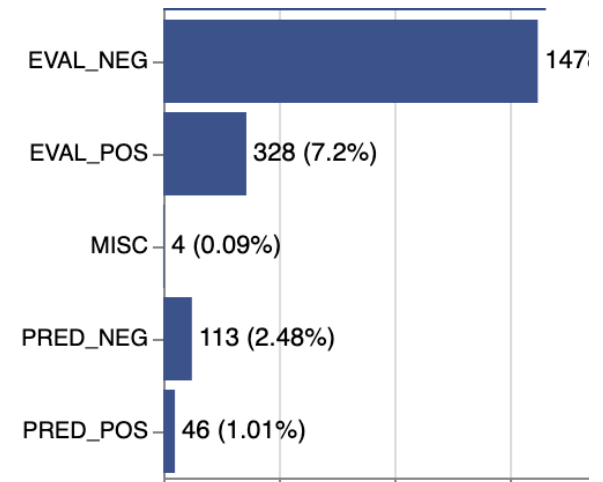
```
...
in mind: to deliver on our price sta
; have little option: they have to d
; how much tightening that might nee
ful transmission belt, or harm, by c
/ to support the markets and economy
ers once again. This is true for hou
ernments, such as the UK\u2019s, are
has been swimming naked.", "meta": {"p
t, "_task_hash": 620770100, "tokens": [{"
id": 1, "ws": true}, {"text": "strength",
: true}, {"text": "the", "start": 20, "end"
": "matters", "start": 31, "end": 38, "id
": "start": 39, "end": 42, "id": 8, "ws": true
53, "end": 57, "id": 10, "ws": true}, {"te
}, {"id": 12, "ws": true}, {"text": "out", "
": true}, {"text": "a", "start": 74, "end"
xt": "US" "start": 83 "end": 85 "id": 1
```

DISPARSA Annotation Report - Co

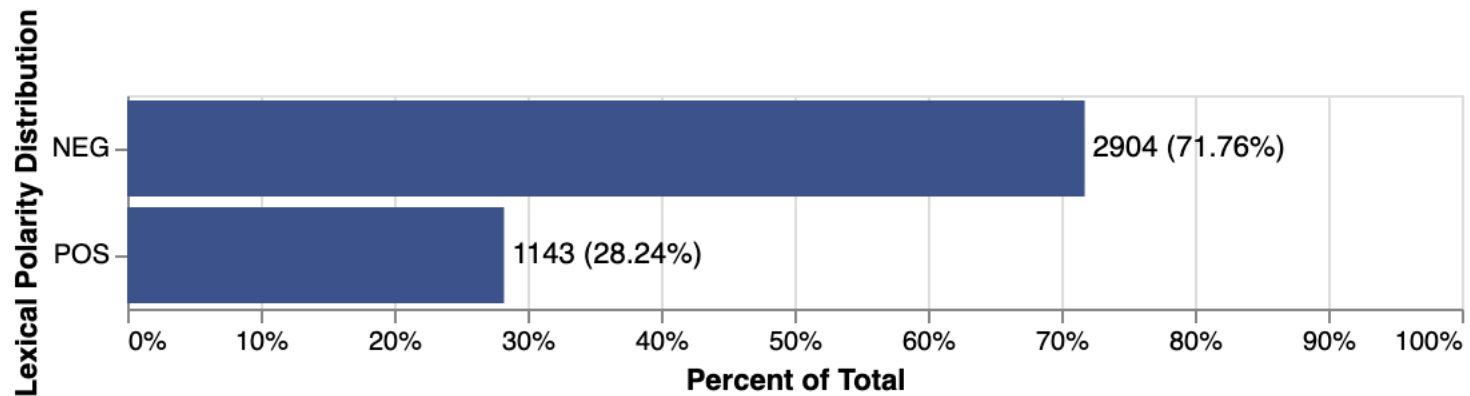
Text 1

- Publisher: FT
- Author: Martin Wolf
- Date: 2022-09-23
- Fdu pattern: ['D', 'D', 'D', 'C', 'D', 'C', 'C', 'D', 'E', 'E', 'A', 'E', 'P', 'E', 'E', 'E', 'E', 'D', 'D', 'E', 'M', 'E', 'C', 'D', 'C', 'D', 'P', 'D', 'E', 'A', 'E', 'E',

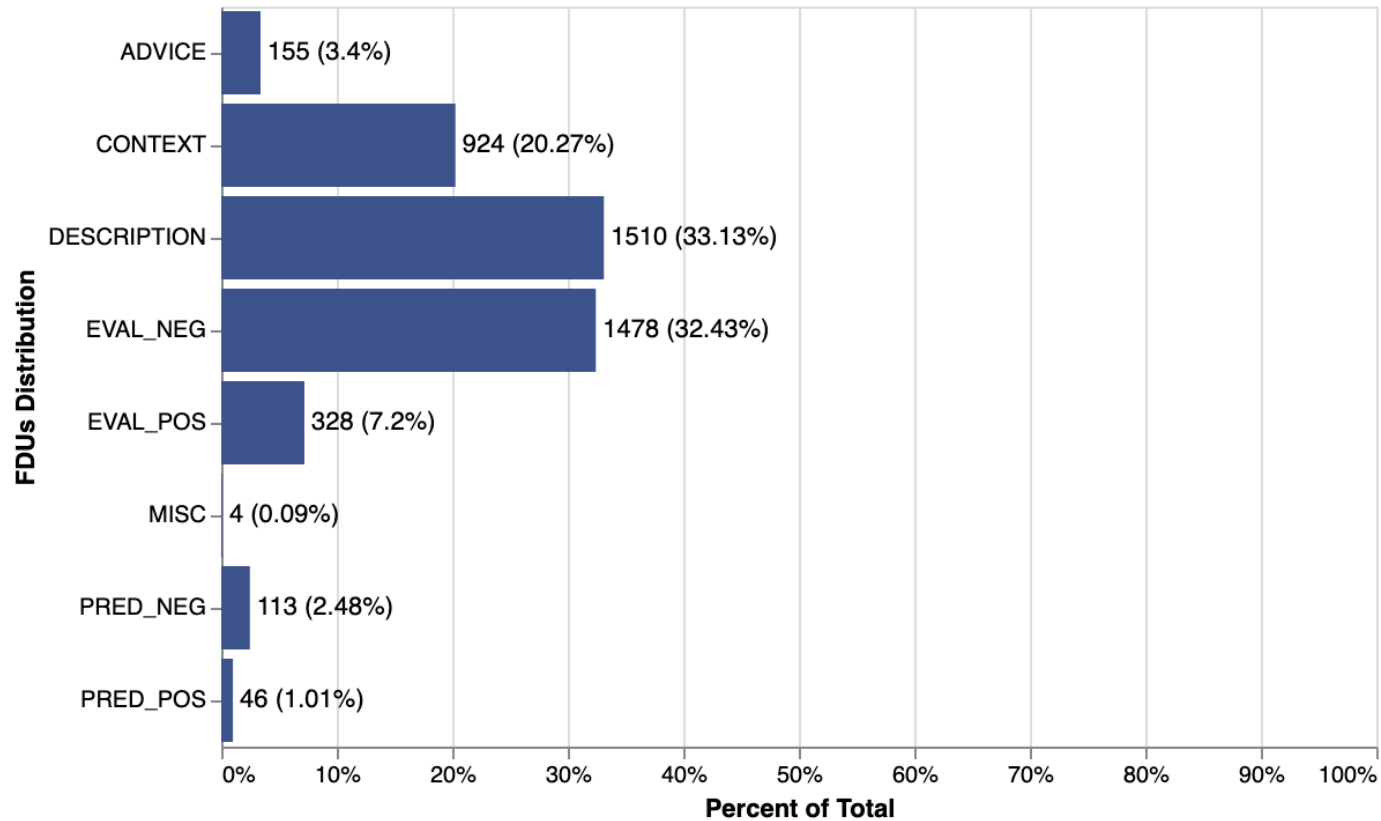
Why the strength of the dollar matters The financial tide is going out as a trouble, the dollar is the world's refuge and strength. This is true even with the crisis of 2007-09. It is true again now. A series of shocks, including high inflation, has hit the dollar. Moreover, this has not been just against the currencies of emerging markets. Meanwhile, the general story of the dollar cycle underlies some specific policy management, proves particularly dangerous when the dollar is strong, in please note. The nominal effective exchange rate of the US dollar appreciated according to JPMorgan estimates. Over the same period, the yen's effective rate fell by 3 per cent. Against the dollar alone, movements are larger: the euro by 16 per cent. The dollar is king of the castle. So why has this happened? The answer is that the world economy has suffered four linked shocks since the pandemic: supply side, in which pent-up (and lopsided) demand hit supply; demand; Russia's invasion of Ukraine, which hit energy, particularly for Europe. TI



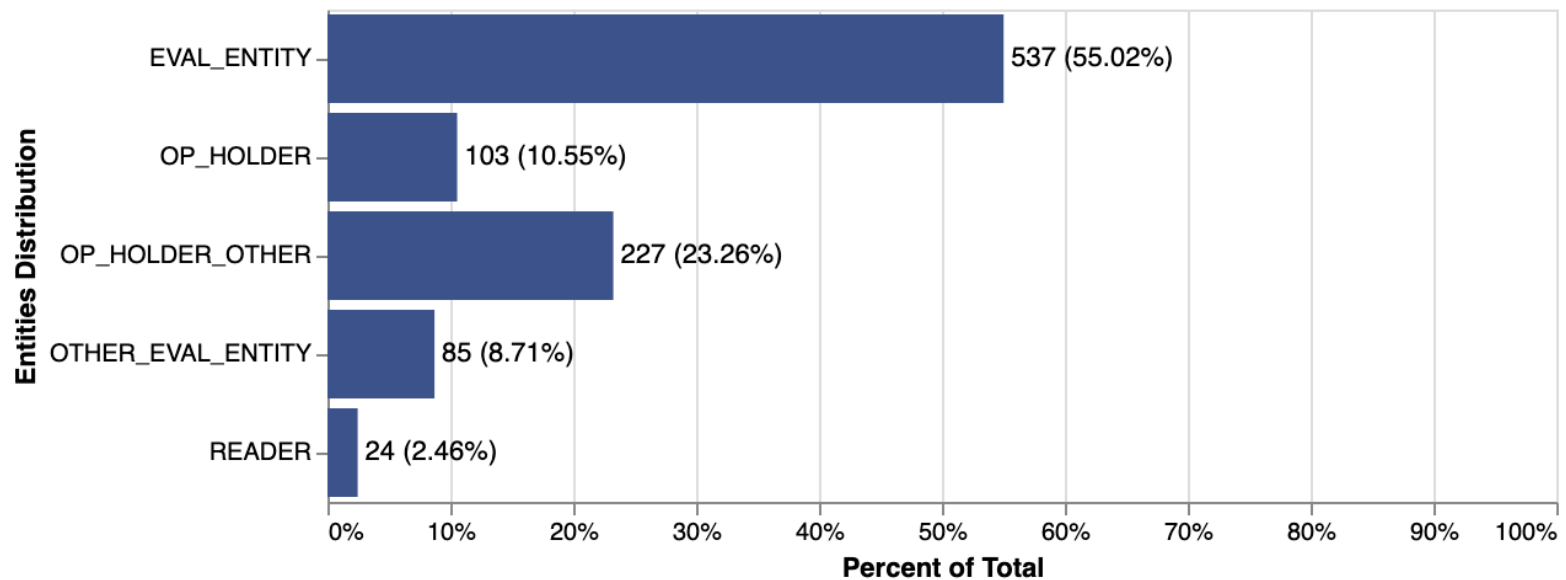
Lexical Polarity Distribution



FDUs Distribution



Entities Distribution



DATA ANALYSIS

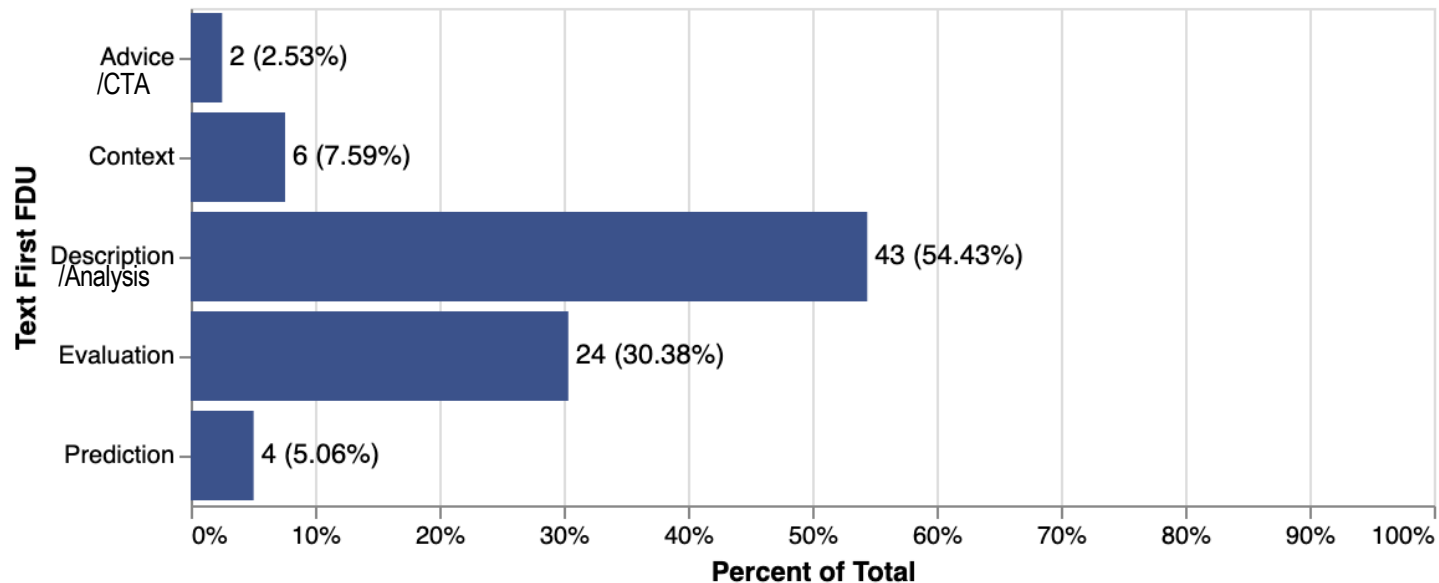
Quantitative Insights



UNIVERSIDAD
DE MÁLAGA

uma.es

Text First FDU: Headline



DATA ANALYSIS

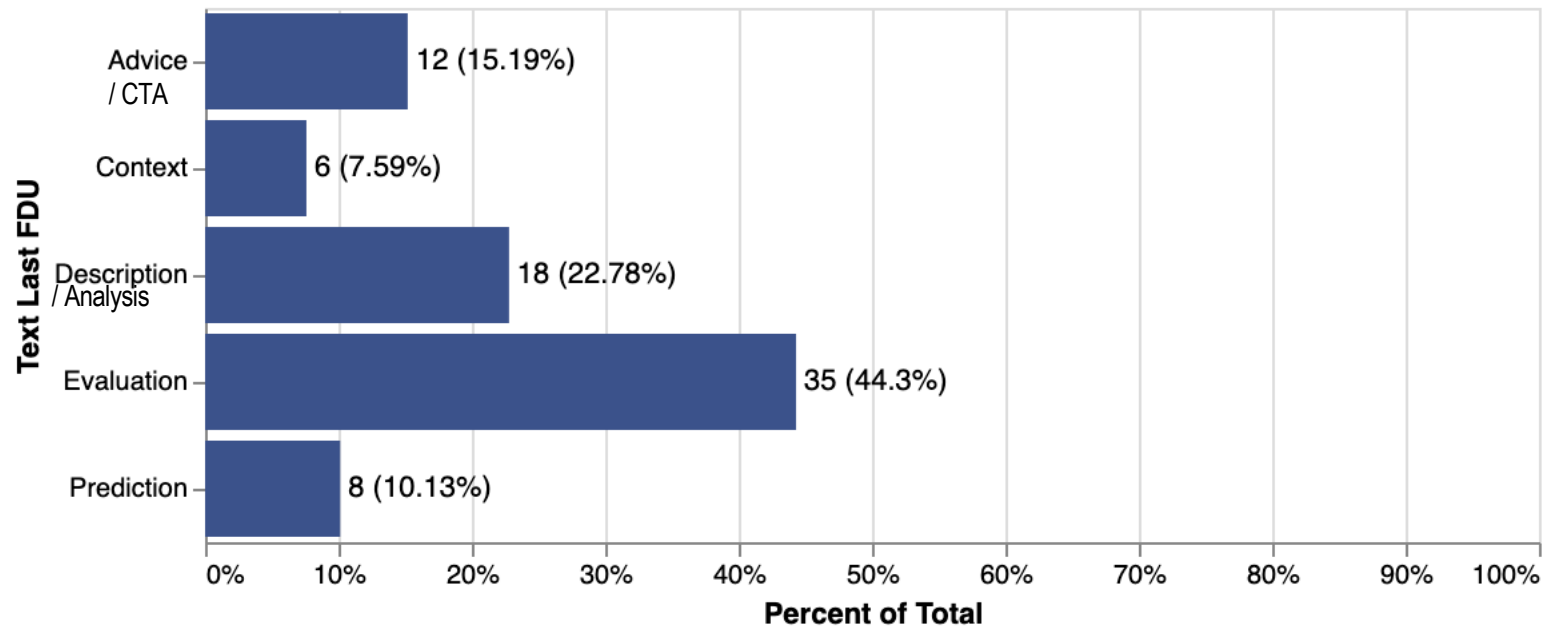
Quantitative Insights



UNIVERSIDAD
DE MÁLAGA

uma.es

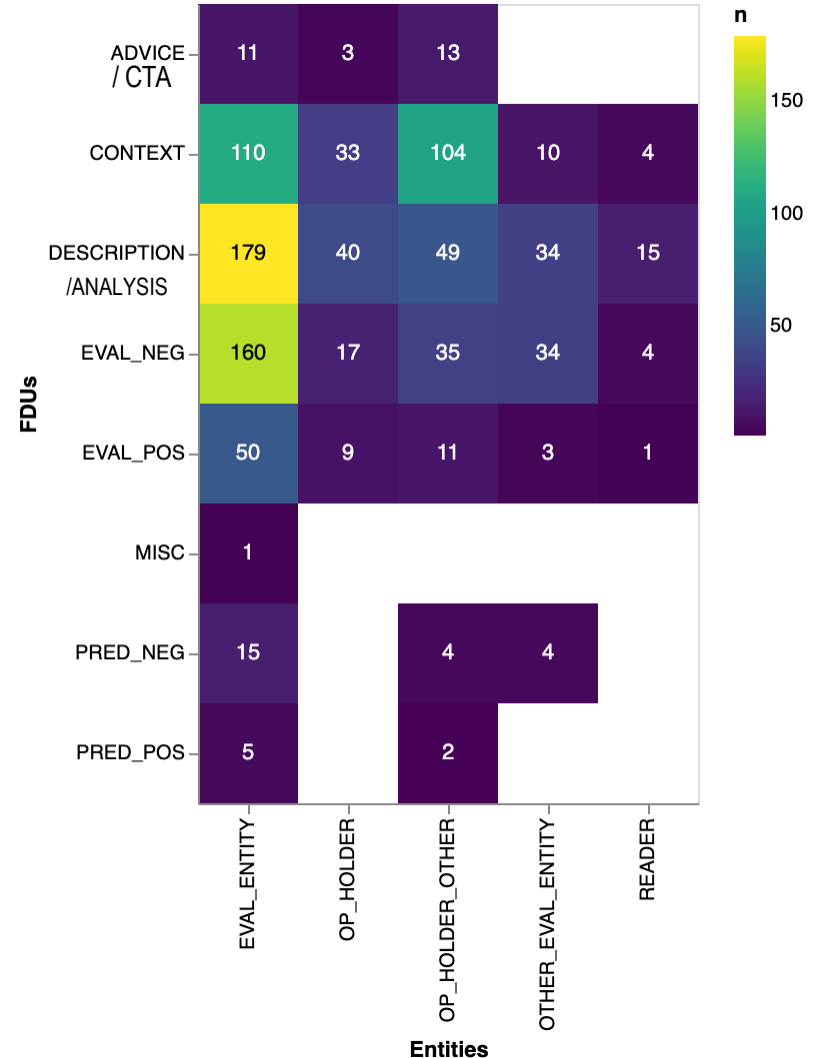
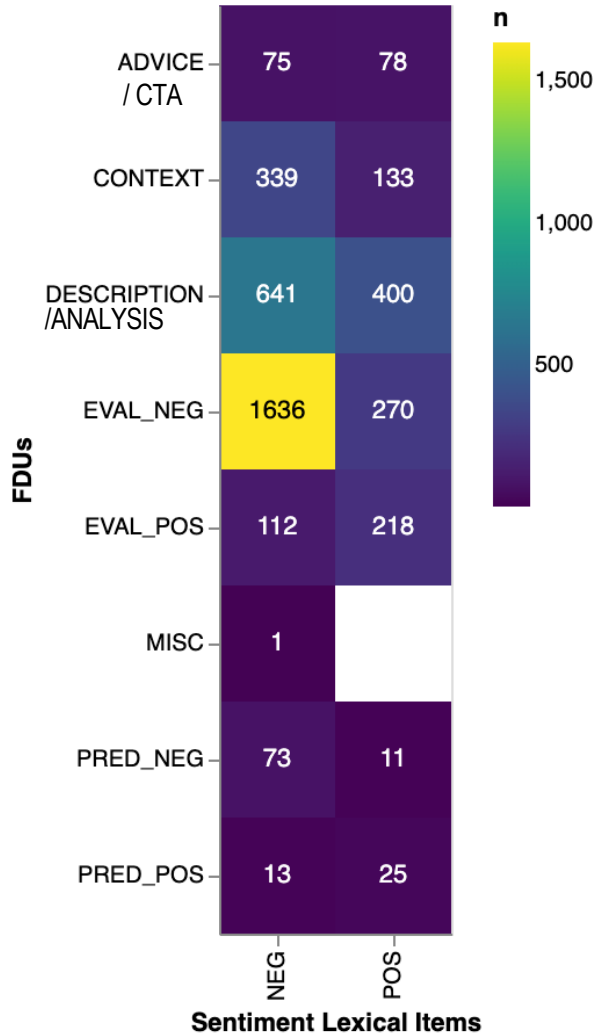
Text Last FDU





DATA ANALYSIS

Quantitative Insights



IRONY

What's my plan for the next two years? I will be happy, healthy and

FDU:CONTEXT

FDU:EVAL_POS

ENT:OP_HOLDE LEX:POS

LEX:POS

successful. What will I do to achieve these things? What are you, a

FDU:CONTEXT

FDU:EVAL_NEG

LEX:POS

Marxist? ↵

(Sort of dry) HUMOR

The reality is that Truss does not have a growth plan. She has a “

FDU: EVAL_NEG

FDU: EVAL_NEG

growth plan” — a magical potion into which she sprinkles the

FDU: EVAL_NEG

LEX: POS

reversal of recent tax increases, freedom for bankers’ bonuses and

lower taxes for the prosperous, says “abracadabra” and suddenly

LEX: POS

trend productivity growth quadruples, conjuring 2.5 per cent annual

LEX: POS

growth. ↵

TOO SHORT FDU_s WITH TWO EVAL_S

Glimmers of light in a terrible year. ↵

FDU:EVAL_POS

LEX:POS

LEX:NEG

DARKER SHADES OF SENTIMENT

Worse may still be to come, possibly even much worse.

LEX:NEG

LEX:NEG

FDU:PRED_NEG

CONCLUSIONS



UNIVERSIDAD
DE MÁLAGA

| uma.es

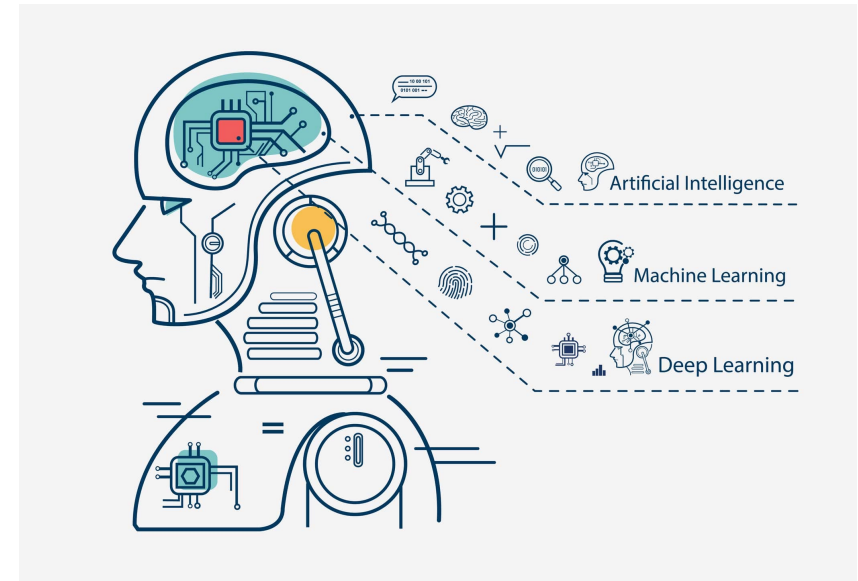
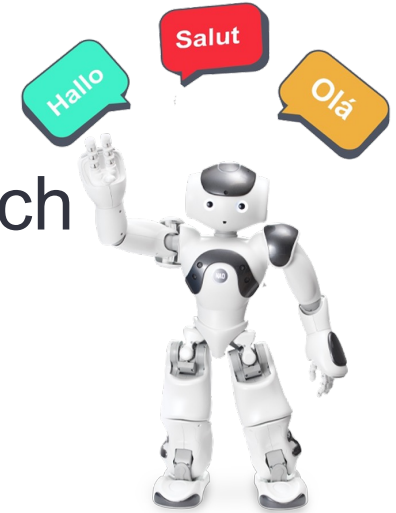
- We have developed a schema for the annotation of Opinion news on Economy.
- It is an original annotated corpus (Open Data!) that fills a research gap.
- It fits opinion economic analysis properly.
- It adjusts to our needs without being too specific (i.e., many tags) nor inefficient (i.e., too little).
- Annotation does not differ too much from other DisparSA corpora.
- Labor-intensive.



(nearly, two more slides)

Results applications (promising!) are bifold:

- Quantitative text-level linguistic research
- FDU detection for Sentiment Analysis
→ Deep Learning models



- Inter-annotator agreement (coming soon!)
- More newspapers – Ideological spectrum
- Contrastive studies.
- Stylistics: Martin Wolf from the FT specifically but we could settle for Krugman's humor.
- Recommendations for Spotify playlists/podcasts while annotating =)





THANKS

Questions?

fernandezcruz@uma.es

irinamt@uma.es



UNIVERSIDAD DE MÁLAGA



UNIVERSITÀ
DEGLI STUDI
DI BERGAMO

- Bednarek, M., & Caple, H. (2014). Why do news values matter? Towards a new methodological framework for analysing news discourse in Critical Discourse Analysis and beyond. *Discourse & Society*, 25(2), 135–158. <https://doi.org/10.1177/0957926513516041>
- Caple, H., & Bednarek, M. (2016). Rethinking news values: What a discursive approach can tell us about the construction of news discourse and news photography. *Journalism*, 17(4), 435–455. <https://doi.org/10.1177/1464884914568078>
- Dafouz-Milne, E. (2008). The pragmatic role of textual and interpersonal metadiscourse markers in the construction and attainment of persuasion: A cross-linguistic study of newspaper discourse. *Journal of Pragmatics*, 40(1), 95–113. <https://doi.org/10.1016/j.pragma.2007.10.003>
- Egbert, J., Wizner, S., Keller, D., Biber, D., McEnery, T., & Baker, P. (2021). Identifying and describing functional discourse units in the BNC Spoken 2014. *Text & Talk*, 41(5–6), 715–737. <https://doi.org/10.1515/text-2020-0053>

- Liu, B. (2015). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press.
- Mann, W. C., & Thompson, S. A. (1987). *Rhetorical Structure Theory: A Theory of Text Organization (Report ISI/RS-87-190)*. Information Sciences Institute.
- Marcu, D. (1997). *The Rhetorical Parsing, Summarization, and Generation of Natural Language Texts [Thesis]*. <http://citeseer.nj.nec.com/24902.html>
- Marcu, D. (1999). Discourse trees are good indicators of importance in texts. In I. Mani & M. Maybury (Eds.), *Advances in Automatic Text Summarization* (pp. 123–136). The MIT Press. <http://www.isi.edu/~marcu/papers/summar-book99.pdf>
- Montani, I., & Honnibal, M. (2018). Prodigy: A new annotation tool for radically efficient machine teaching. *Artificial Intelligence*, to appear.
- Moreno Ortiz, A. (2017). Lingmotif: A user-focused sentiment analysis tool. *Procesamiento Del Lenguaje Natural*, 58, 133–140.
- Moreno-Ortiz, A., & Perez-Hernandez, C. (2018). Lingmotif-lex: A wide-coverage, state-of-the-art lexicon for sentiment analysis. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 2653–2659.