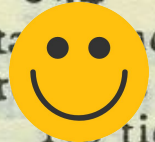


624

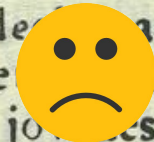
EL CENSOR.

acciones propias, quien recur-
ajenas.

...tienes armas... Mas quiero no
tenerlas, que dar motivo á que pre-
gunten por qué razon las tengo. La
verdadera gloria consiste en merecer
la gloria.



El ardor con que Caton iba á
proseguir en su vehemente de-
cision se interrumpió de repe-
los gritos de una turba de jó-
que no se que novedad inesperada
atrahia al foro. Vuelve la cabeza el
venerable anciano, y vé ¡ó dolor! ¡ó
oprobio! vé á su nieta, á la hija de



LLMs for NLP: Sentiment Analysis

Lucija Krušić & Martina Scholger

Department for Digital Humanities, University of Graz

DHd-AG Angewandte Generative KI in den Digitalen Geisteswissenschaften (AGKI-DH)

Outline

- Sentiment Analysis for historical research
- Comparison of different methods: dictionary, machine learning, generative AI
- Evaluation according to a manually annotated gold standard
- Corpus-specific challenges
- Future steps

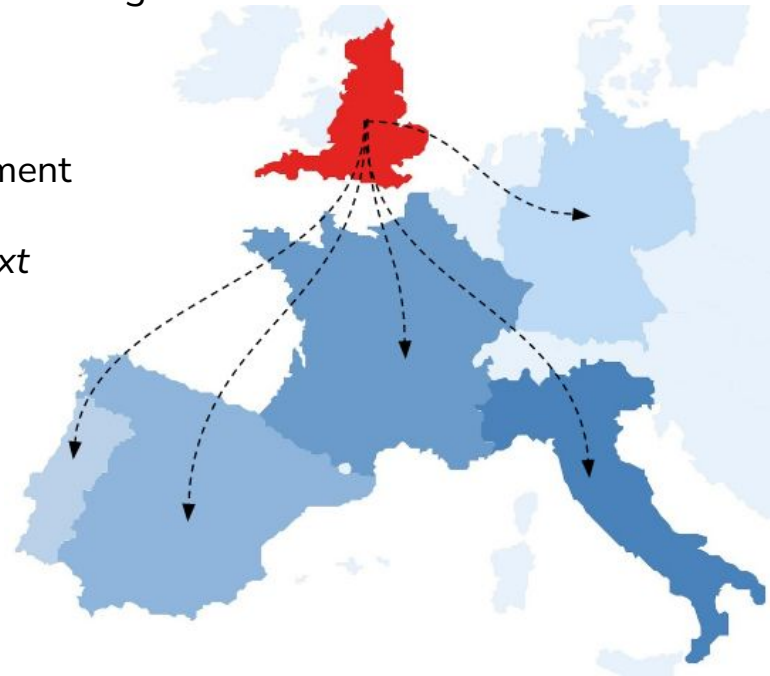
Goals

- How do authors of 18th-century Spanish periodicals utilize tone to educate their readership
- Does the tone differ depending on the use of specific narrative forms?

Research corpus

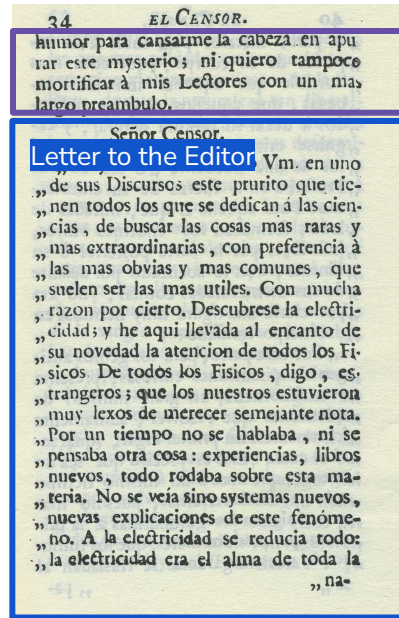
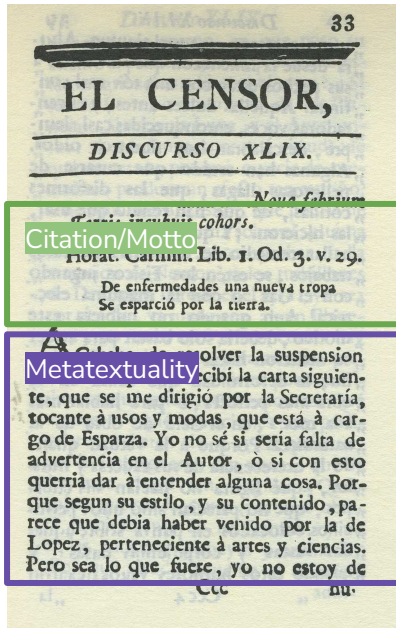
The Spectator Press

- Journalistic-literary genre of the 18th century, created in England
- Imitations and translations in Europe
- Propagating social norms and values of the Enlightenment
- Digital Edition *The Spectators in the International Context* (Ertler et al. 2011-2022) and its analysis in *Distant Spectators* (Scholger et al. 2021-2022)
- Spanish, French, Italian, Portuguese, German, English



Research corpus: The Spectators

TEI Encoding of narrative forms



<p>

<milestone ana="#MT" spanTo="#FR.4" unit="form"/>

Acababa de resolver la suspension de mi obra quando recibí la carta siguiente, que se me dirigió por la Secretaría, tocante à usos y modas, que está à cargo de Esparza. Yo no sé si sería falta de advertencia en el Autor, ò si con esto querria dar à entender alguna cosa. Porque segun su estilo, y su contenido, parece que debia haber venido por la de Lopez, perteneciente à artes y ciencias. Pero sea lo que fuere, yo no estoy de

<pb facs="#T-III 034" n="34"/> humor para cansarme la cabeza en apurar este mysterio; ni quiero tampoco mortificar à mis Lectores con un mas largo preambulo.

<anchor xml:id="FR.4"/>

</p>



Documentation

Periodicals

Volltextsuche



Citation: Anonym (García de Cañuelo, Luis; Pereira, Luis Marcelino) (Ed.): "Discurso XLIX", in: *El Censor*, Vol.3/049 (1783), pp. , edited in: Ertler, Klaus-Dieter/Hobisch, Elisabeth (Ed.): Spectators. Digitale Edition, Graz 2011- . hdl.handle.net/11471/513.20.395 [last accessed: 27.09.2023].

► Level 1

Discurso XLIX

► Citation/Motto Noua februm

Terris incubuit cohors.

Horat. Carmin. Lib. 1. Od. 3. v. 29.

De enfermedades una nueva tropa

Se esparció por la tierra. ◀ Citation/Motto

► Level 2 ► Metatextuality Acababa de resolver la suspension de mi obra quando recibí la carta siguiente, que se me dirigió por la Secretaría, tocante à usos y modas, que está à cargo de Esparza. Yo no sé si sería falta de advertencia en el Autor, ò si con esto querria dar à entender alguna cosa. Porque segun su estilo, y su contenido, parece que debia haber venido por la de Lopez, perteneciente à artes y ciencias. Pero sea lo que fuere, yo no estoy de humor para cansarme la cabeza en apurar este mysterio; ni quiero tampoco mortificar à mis Lectores con un mas largo preambulo. ◀ Metatextuality

► Level 3 ► Letter/Letter to the editor "Señor Censor.

Muy Señor mio: notó Vm. en uno de sus Discursos este prurito que tienen todos los que se ocupan de las ciencias, de buscar las cosas mas raras y mas extraordinarias, con preferencia à las mas comunes, que suelen ser las mas utiles. Con mucha razon por cierto. Descubrese he aquí llevada al encanto de su novedad la atencion de todos los Fisicos. De todos los Fisicos, digo, estrangeros; que los nuestros estuvieron muy lexos de merecer semejante nota. Por un tiempo no se hablaba, ni se pensaba otra cosa: experiencias, libros nuevos, todo rodaba sobre esta materia. No se veía

Original view

Analytical view

Narrative levels of representation ⓘ

Level 1

Level 2

Level 3

Level 4

Level 5

Narrative forms ⓘ

General account

Satirisches Erzählen

Metatextuality

Citation/Motto

Letter/Letter to the editor

↓ Download

TEI

TXT

Analytical view

<https://gams.uni-graz.at/spectators>

Sentiment analysis

WHAT?

polarity



emotions



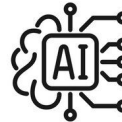
HOW?



Lexicon (dictionary) based



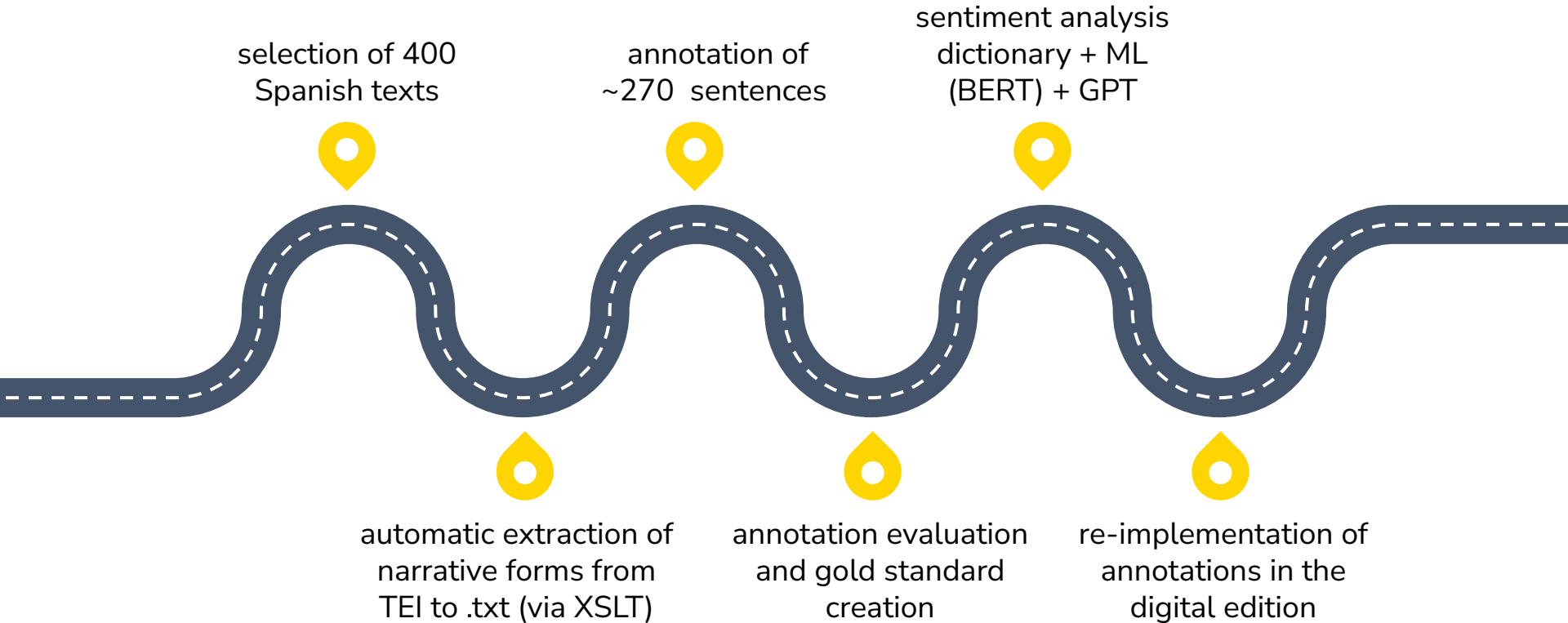
Hybrid & ML-based



Generative AI

Research corpus

Workflow



Research corpus

Analysis of narrative forms

Narrative form (NF)	Number of NF	Number of tokens
metatextuality (MT)	580	267,010
example (EX)	500	58,801
general account (AE)	422	389,663
letters (LB)	408	506,209
citation (ZM)	467	13,073
heteroportrait (FP)	326	84,767
dialogue (D)	159	131,136
selfportrait (SP)	87	52,167
satire (S)	44	63,685
dream (TR)	27	43,916
fable (F)	19	6,507
allegory (AL)	8	16,186
utopia (UT)	6	6,968

Sentiment analysis

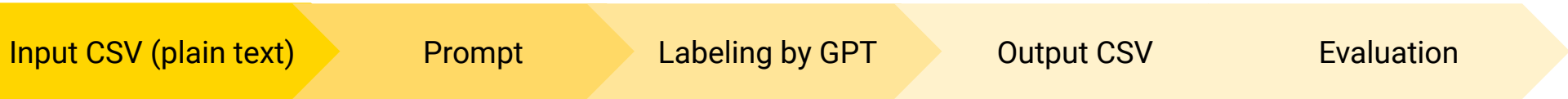
Machine learning

- a transformer-based open-source Python library for Spanish SA, *pysentimiento* (Pérez et al, 2021)
- offers Sentiment Analysis, Emotion Analysis, Hate Speech Detection, Named Entity Recognition and POS Tagging in Spanish and English
- **BERT-based** (RoBERTuito) trained on the TASS Dataset in Spanish

Sentiment analysis

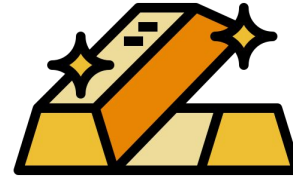
Generative AI

- Experiments on data labeling with GPT-4 and GPT-4o (using the API), temperature 0.2
- Following [Whitaker, J. \(2023\)](#): Bootstrapping Labels with GPT-4. A cost-effective approach to data labeling.



Gold standard for evaluation

- benchmark for evaluation of SA methods
- created by majority vote
- 4 annotators: 227 (out of 271) sentences



Annotation study

- Method: survey (LimeSurvey)
- 3 attempts: 2, 3 and 4 expert annotators
- instances - representative of the corpus, balanced

In this survey you will be provided with three sentences per question.

The first and the third sentence will mark the context (previous and following sentence). The second, middle sentence is the one that needs to be annotated and will be marked in **bold**. There are a few cases where either the previous sentence or the following sentence is missing.

Please annotate the sentences in **bold** with **BOTH** a polarity **AND** an emotion option. The polarity is whether you find the sentence to express a positive, neutral or a negative sentiment.

Furthermore, please associate one of the listed emotions with the sentence. If you believe that the sentence expresses no emotion, check the "no emotion" option.

*

Previous sentence:

Pero ella, conociendo el riesgo que había en verle, y escucharle, llena de indignación, y colera, le negó su presencia.

To annotate:

Mi pobre Vecino, poseído de una furiosa frenesí, enagenado de sí mismo, y sofocado de rabia, se enfurecía, y maldecía su fortuna, y precipitada audacia.

Following sentence:

Hasta que volviendo en sí, se halló confuso de sus extravagancias, è intentò, arrepentido, trocar su enojo en una sumisión profunda; pero nada sirvió para el caso.

Check all that apply

positive

joy

surprise

sadness

neutral

trust

anger

fear

negative

anticipation

disgust

no emotion

Previous

Next

Sentiment annotation

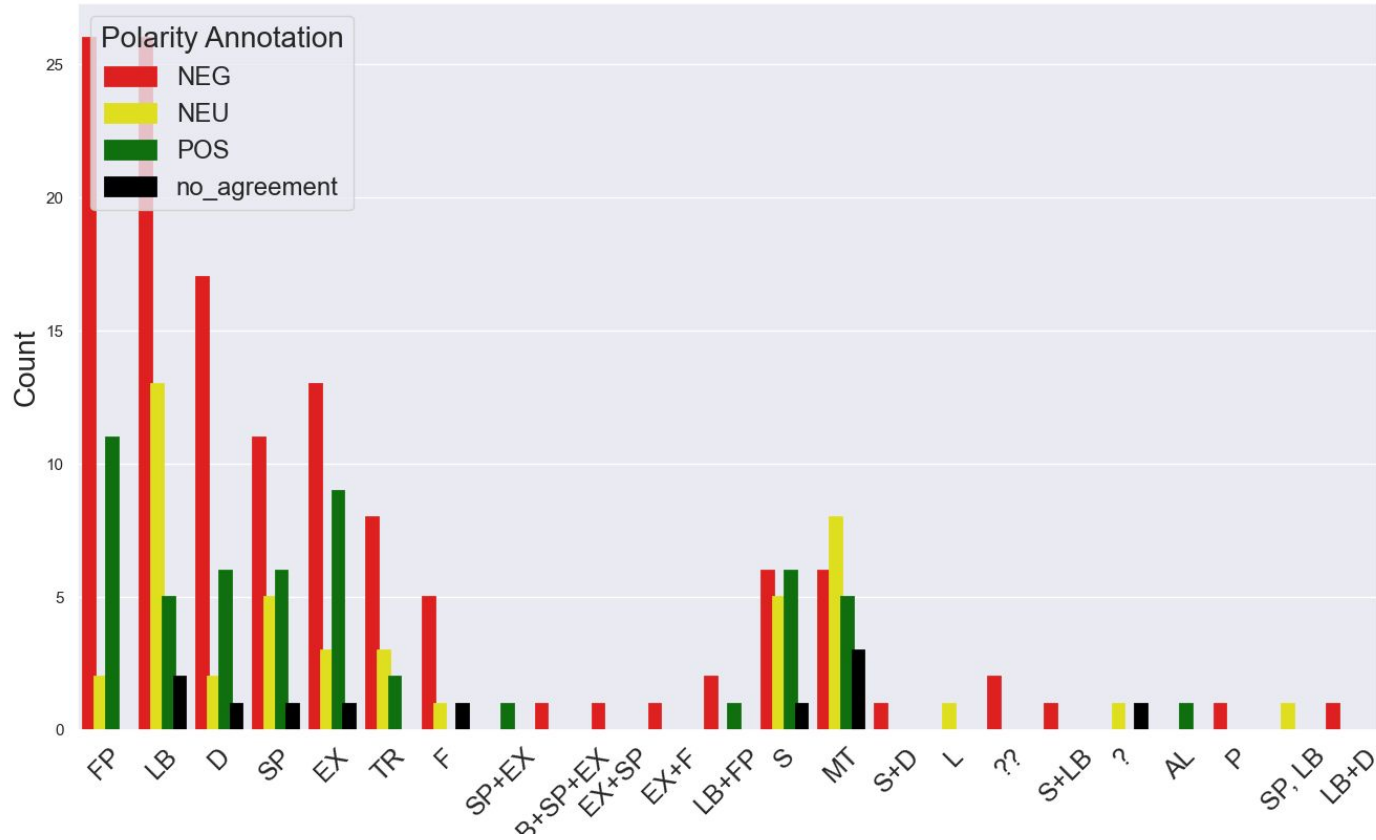
Intra-rater agreement

Nr. annotators	2	3	4
Cohen's Kappa - polarity	0.31	0.41	0.46
Cohen's Kappa - emotion	0.2	0.14	0

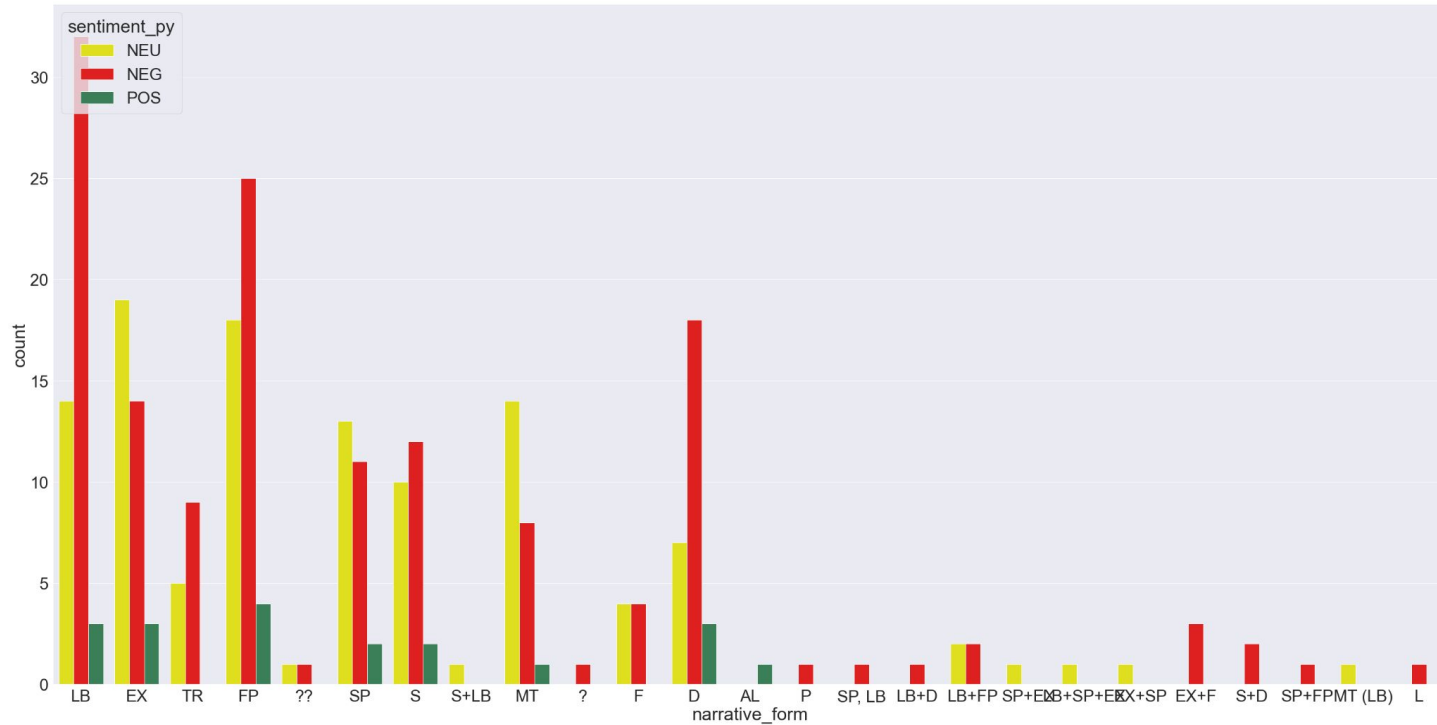
Cohen's kappa statistic	Strength of agreement
< 0.00	Poor
0.00-0.20	Slight
0.20-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost perfect

McHugh M.L. 2012

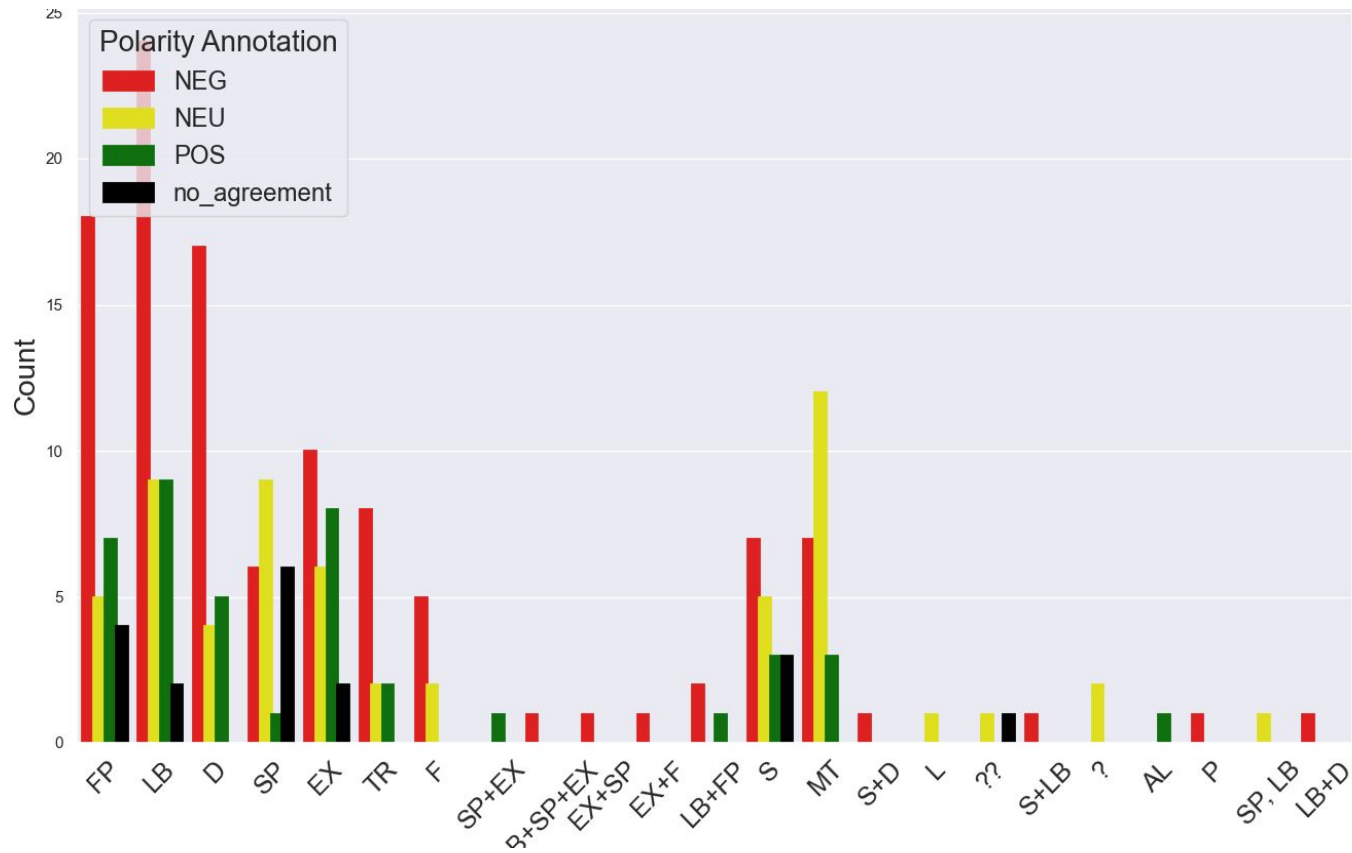
Gold standard - Polarity - Human annotators



Polarity ML method (pysentimiento)



Polarity GPT-4



Simple prompt for GPT 4 and GPT 4o

You are an expert in sentiment analysis of 18th-century Spanish literature.

Classify the sentiment of each sentence from the input `{input_text}` as 'positive', 'negative', or 'neutral'. Add the sentiment of each sentence in a new column 'polarity'.

Advanced prompt for GPT 4 and GPT 4o

You are an expert in sentiment analysis of 18th-century Spanish literature.

Persona

Carefully analyse the input and classify each text {input_text} as either (positive, neutral, negative).

Task

Address specific challenges:

****epoch-specific meanings****: interpret words and phrases as they were used in the 18th century

****non-standardized orthography****: manage variations in spelling and grammar typical of the period.

****opinion shifters**** identify and interpret sarcasm, satire, irony, negations, and idiomatic expressions. For instance, sarcasm might reverse the apparent sentiment of a sentence.

****Record your classification in a new column 'polarity'.**

**Guidelines
+ Context**

This task is crucial for understanding the socio-historical context and emotional undertones of the texts.

Take a deep breath and work on this problem step-by-step. This is very important to my career!

Read the instructions again before you continue with annotating the next line.

Emotional prompting

Sentiment evaluation

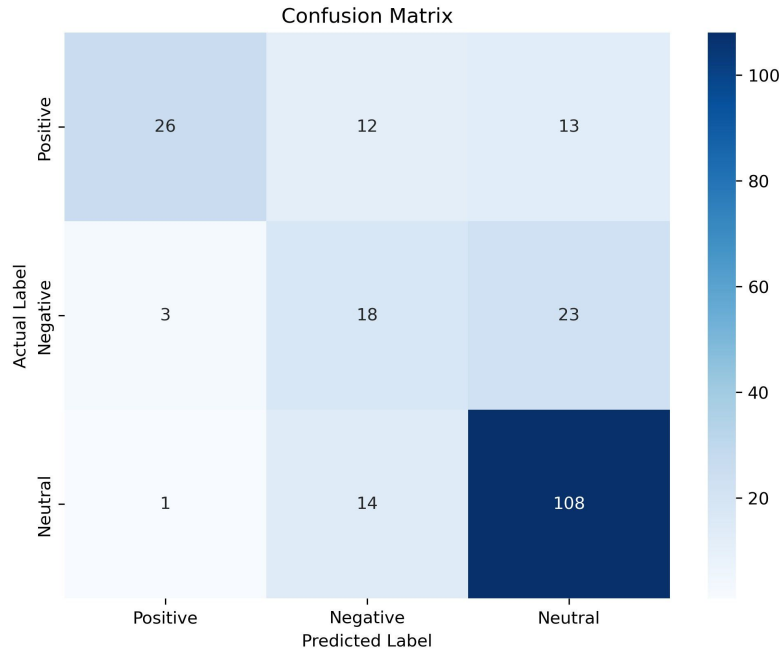
Lexicon - pysentimiento - GPT 4 (April 2024)

Method	Accuracy	F-1 (macro) score	unknown/mixed
Lexicon corrected (Koncar et al. 2022)	0.46	0.44	0
<i>pysentimiento</i>	0.70	0.62	0
GPT 4 (simple prompt)	0.38	0.39	94
GPT 4 (advanced prompt)	0.86	0.50	7

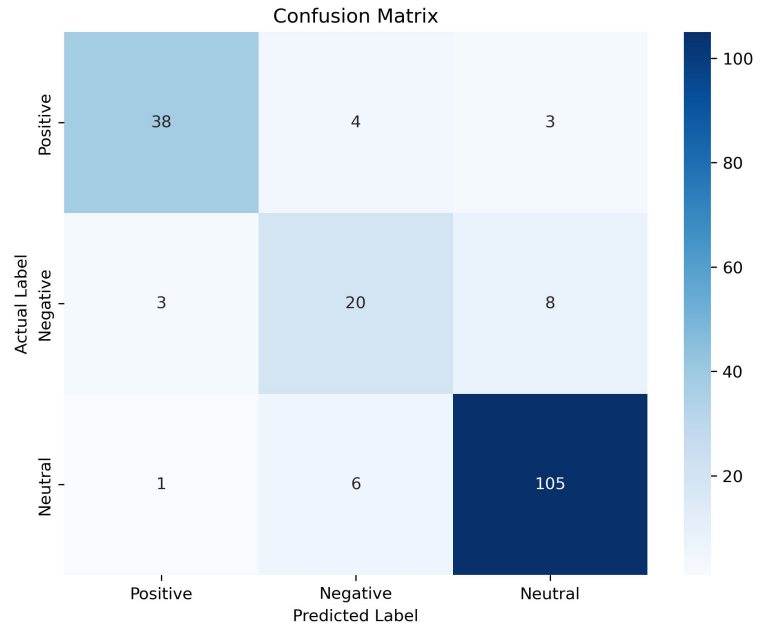
Sentiment evaluation

Confusion matrices

pysentimiento



GTP-4



Sentiment evaluation

Lexicon, pysentimiento, GPT 4 and GPT 4o (Sept 2024)

Method	Accuracy	F-1 (macro) score	unknown
Lexicon corrected	0.46	0.44	0
<i>pysentimiento</i>	0.73 (0.70)	0.65 (0.62)	0
GPT 4 (simple)	0.82 (0.38)	0.77 (0.39)	0 (94)
GPT 4 (advanced)	0.78 (0.86)	0.53 (0.50)	4 (7)
GPT 4o (simple)	0.82	0.77	0
GPT 4o (advanced)	0.64	0.42	4

Corpus-Specific challenges

- epoch-specific meanings (el cortejo)
- orthographic fluidity (muger → muj^rer)
- semantic ambiguity due to orthography
- historical meaning vs. contemporary (el coche)
- Opinion shifters
 - sarcasm, irony, satire
 - negations
 - idiomatic expressions (bajarse del burro → lit.: to get off the donkey → sense: to admit you were wrong)



Human evaluation of GPT results

by E. Hobisch and Y. Völkl

Example 1: The following sentence was rated as positive by *pysentimiento*, but as negative by GPT and human annotators:

“Todo esto llevará con paciencia, si su asèo, y prolixa compostura se ordenára â conplacermè; pero es muy â el contrario, porque su merced es Cortejo, y hace tanta vanidad de serlo, que quando està de humor me refiere las gracias de su querida, me pondera sus chistes, y alaba su hermosura: vea Vm. què refresco para quien se halla atabardillada de Zelos.”

Translation: I would bear all this with patience if his neatness and prolix composure were ordered to please me; but it is quite the contrary, because this Mister is a Cortejo [impossible to translate], and makes so much vanity of being one, that when he is in the mood he refers to me the graces of his dearest, and praises me her jokes, and praises her beauty: you see what a refreshment for one who is sick of jealousy.

positive negative sarcasm

Human evaluation of GPT results

by E. Hobisch and Y. Völkl

Example 2: The next example is a two-part sentence in which each part expresses a different sentiment. It is rated negative by *pysentimiento* and positive by GPT and human annotators:

“Convencerá tal vez nuestro espíritu; pero poco importa: no gana, no somete nuestro corazón; y à pesar de sus raciocinios artificiosos, tenemos bastante fuerza para resistirle”

Translation: He will perhaps convince our spirit; but it matters little: he does not win, he does not subdue our heart; and in spite of his contrived reasonings, we have strength enough to resist him.

positive negative shifter

Next steps

- Investigate scope of prompt (simple vs. advanced vs. something in between)
- Variation in temperature
- Provide more context
 - historical vocabulary (e.g. Cortejo)
 - gold standard of annotated sentences for analyzing the whole corpus
- Investigate other models => Llama 3

Thank you for you attention!

Contact

- Lucija Krušić | lucija.krusic@uni-graz.at
- Martina Scholger | martina.scholger@uni-graz.at

GitHub repo

<https://github.com/lucijakrusic/SentientSpectators>

Thanks to our project partners:

- Bernhard Geiger, Know-Center Graz
- Elisabeth Hobisch, Technical University Graz
- Yvonne Völkl, Technical University Graz

Sentiment annotation:

- Klaus-Dieter Ertler
- Elisabeth Hobisch
- Yvonne Völkl
- Pascal Striedner

References

- Bsharat, Sondos Mahmoud, Aidar Myrzakhan, and Zhiqiang Shen. “Principled Instructions Are All You Need for Questioning LLaMA-1/2, GPT-3.5/4.” arXiv, December 26, 2023. <https://doi.org/10.48550/arXiv.2312.16171>.
- Ertler, K.-D., Fuchs A., Fischer-Pernkopf M., Hobisch E., Scholger M. & Völkl Y. (2011-2021). The Spectators in the International Context. <https://gams.uni-graz.at/spectators>.
- Koncar, P., Geiger, B. C.; Glatz, C.; Hobisch, E., Sarić, S., Scholger, M., Völkl, Y., Helic, D. (2022): A Sentiment Analysis Tool Chain for 18th Century Periodicals. In: Manuel Burghardt, Lisa Dieckmann, Timo Steyer, Peer Trilcke, Niels-Oliver Walkowski, Joëlle Weis, Ulrike Wuttke (Eds.): Fabrikation von Erkenntnis. Experimente in den Digital Humanities. Luxembourg. Zeitschrift für digitale Geisteswissenschaften und Melusina Press. 2022. DOI: <https://doi.org/10.26298/ezpg-wk34>.
- Liu, B. (2012) Sentiment Analysis and Opinion Mining (Synthesis Lectures on Human Language Technologies). Morgan & Claypool Publishers, Vermont, Australia.
- McHugh ML.(2012) Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012;22(3):276-82. PMID: 23092060; PMCID: PMC3900052.
- Pérez, J. M., Giudici, J. C., & Luque, F. (2021). pysentimiento: A Python Toolkit for Sentiment Analysis and Social NLP tasks. ArXiv:2106.09462 [Cs]. <http://arxiv.org/abs/2106.09462>.

References

- Schmidt, T., Burghardt, M. & Dennerlein, K. (2021). Using Deep Learning for Emotion Analysis of 18th and 19th Century German Plays. In: Manuel Burghardt, Lisa Dieckmann, Timo Steyer, Peer Trilcke, Niels-Oliver Walkowski, Joëlle Weis, Ulrike Wuttke (Eds.): Fabrikation von Erkenntnis. Experimente in den Digital Humanities. Luxembourg. Melusina Press. DOI: <https://doi.org/10.26298/melusina.8f8w-y749-udlf>.
- Scholger, M., Geiger, B., Hobisch, E., Koncar, P., Sarić, S., Völkl, Y.; Glatz, C. (2019-2021): *Distant Spectators (DiSpecs)*. *Distant Reading for Periodicals of the Enlightenment*. <https://gams.uni-graz.at/dispecs>.
- Sidorov, G., Miranda-Jiménez, S., Viveros-Jiménez, F., Gelbukh, A., Castro-Sánchez, N., Velásquez, F., Díaz-Rangel, I., Suárez-Guerra, S., Treviño, A., & Gordon, J. (2012). Empirical study of opinion mining in Spanish tweets. In LNAI 7630 (pp. 1-14).
- Sprugnoli, R., Passarotti, M., Corbetta, D. & Peverelli, A. (2020). Odi et Amo. Creating, Evaluating and Extending Sentiment Lexicons for Latin. In Proceedings of the 12th Language Resources and Evaluation Conference, 3078-3086.
- Whitaker, J. (2023): Bootstrapping Labels with GPT-4. A cost-effective approach to data labeling. <https://towardsdatascience.com/bootstrapping-labels-with-gpt-4-8dc85ab5026d>.