

A Corpus of Misunderstood Irony on Turkish Social Media

Çağrı Çöltekin, Güliz Güneş

University of Tübingen

cagri.coeltekin@uni-tuebingen.de, gueliz.guenes@uni-tuebingen.de

Abstract

We present a new Turkish social media corpus annotated for verbal irony. The ironic post candidates are identified by a distant supervision method relying on reports of misunderstood irony in social media platforms. The data collected through this method, as well as irony-tagged posts and a random sample of posts are annotated by three annotators, resulting in a corpus of 3 000 tweets with high quality annotations that may be useful for linguistic analysis as well as for training automatic irony detection systems or testing irony understanding of large language models. Since irony interpretation typically involves context, our dataset also includes the preceding conversational context of the potentially ironic expression. Besides the description of the corpus and the annotation process, this paper presents an analysis of the corpus. Our findings indicate that relying on distant supervision alone may result in suboptimal labels for irony/sarcasm corpora. We also investigate the usefulness of context for the annotators in identifying irony.

Keywords: irony, sarcasm, corpus, Turkish, context

1. Introduction

Figurative language, where the literal meaning of a linguistic expression diverges from the intended meaning, poses a challenge for both humans and natural language processing (NLP) systems. Among figures of speech, irony probably poses the most difficult case, since it is not as conventionalized as the other figures of speech (e.g., idioms or metaphors), and correct interpretation of an ironic expression typically requires more contextual knowledge.

Despite the need for clear contextual information, most available datasets for irony and sarcasm do not contain any information regarding the context of the potentially ironic utterance.¹ Only a few existing datasets annotated for irony include context. These involve almost exclusively the conversational context surrounding the ironic utterances or social media posts. As well as the conversational context, recognizing irony also depends on the non-linguistic context shared by the participants of the conversation. Successful identification of ironic expressions requires shared knowledge of events, and cultural references. The linguistic and non-linguistic cues for irony detection are also highly dependent on language and culture. On the other hand, like many other linguistic resources, most datasets created for automatic irony processing are for English. To our knowledge, there are only a few non-English datasets with irony annotations (see Section 2). For automatic irony detection and quantitative studies on irony, there is a clear need for datasets in other

languages.

Another common trend in irony datasets is the use of ‘distant supervision’, or opportunistic annotation methods where labels are extracted automatically from the environment the data is collected from. The most common approach for building datasets for irony detection is based on (hash)tags used in social media, for example, #irony, #sarcasm or #not on Twitter/X, and /s on Reddit. Such approaches to annotation allow much larger datasets compared to human annotated corpora, and were found to be useful for automatic irony identification in the past. However, to the best of our knowledge, these annotation practices are not carefully validated, and it is often unclear whether the annotations obtained automatically from users correspond to a particular definition of irony. Furthermore, relatively few irony datasets are annotated manually, and almost all of the manually annotated datasets are released with a single (aggregated) label per data instance, even when each data instance is labeled by multiple annotators. For highly subjective annotation tasks like irony annotation, this is often not desirable, and may reduce the value of the data for studies where perspectives of individual annotators are important.

In this paper, we introduce a Turkish social media dataset annotated manually by multiple annotators for irony. We follow a distant supervision method similar to Shmueli et al. (2020) to collect our data from Twitter/X. The method makes use of reports of misunderstood irony in the social media conversations. For example, a post stating ‘*She is being sarcastic*’, is likely to point to an earlier ironic post misunderstood by a different social media user. By tracing the earlier posts in the same

¹In this study, we consider sarcasm as a subset of irony, and we do not attempt to distinguish sarcastic and non-sarcastic irony.

thread of conversation, it is possible to find a large number of (potentially) ironic posts in a fast and efficient way. This method capitalizes on the role of contextual information in detecting irony (going beyond single individual posts) and helps build an irony dataset that is composed of posts that are not necessarily signposted with explicit tags, but they are signposted by (other) social media users, who are part of the ongoing conversation and have more common ground knowledge for the correct interpretation. However, like any other automatic means of obtaining labels, the method is susceptible to false positives. Although we use the method to obtain potentially ironic tweets, we also annotate a substantial part of the tweets manually, resulting in a richer and higher quality dataset compared to the approaches that rely purely on distant supervision. As well as providing a linguistic resource that can be used for building automatic irony identification systems, our aim is to also understand the annotator behavior, as well as the conditions and (lack of) cues that lead to (mis)understanding ironic statements. Our corpus contains annotated social media posts with and without conversational context. To our knowledge, our corpus is the first irony corpus of Turkish that includes context and multiple annotations, as well as being the largest corpus of all existing Turkish irony corpora to date.

In the remainder of the paper, we provide a brief survey of some of the major irony datasets (Section 2); we describe our the data collection (Section 3) and annotation (Section 4) method, and discuss the proposed methodology quantitatively and qualitatively based on the corpus created in this study in Section 5 before concluding (Section 6).

2. Related Work

Irony and sarcasm have been considered one of the most challenging aspects of natural language processing since early work in the field. As a result, there is a wide range of datasets with different end goals, and using different approaches. Although it is difficult to do justice to the depth and breadth of the irony datasets created to date, we provide a summary of example datasets in Table 1. Besides corpora that annotate natural occurrences of irony and sarcasm, we also include some examples from satire, which can in principle be useful for supplementing irony detection datasets. The clear, and non-surprising, observation is that the approaches that rely on automatic methods often produce much larger datasets, while manual annotation results in considerably smaller datasets. Most corpora listed in Table 1 do not include context information. Despite some recent work introducing automatically generated irony corpora for other languages (e.g., Đoković and Robnik-

Šikonja, 2024; Kim et al., 2024), most corpora annotated for irony are for English.

The most closely related study to our current study is a small Turkish social media corpus annotated for irony introduced by Ozturk et al. (2021). The corpus contains 300 tweets with irony, and 300 without. Although the tweets are manually annotated, the selection of tweets is not clear, and no context is provided in the corpus. Another work on Turkish with a dataset collected from satirical news is Ozturk et al. (2025). The labels (satirical, not-satirical) in this work is based on where the text appeared (a popular satirical news site, or online newspapers).

A majority of the data collected for irony processing involves searching for (hash)tags, restricting the social media posts to certain topics (e.g., politics) where expectations for ironic expressions is higher. Even when the studies include manual annotation, since ironic expressions are not common in most contexts, a pre-selection based on these methods reduces the annotation efforts significantly. However, this causes a biased selection of source material, which will inevitably affect the irony detection systems, or quantitative studies of irony on these datasets. Another concern regarding datasets generated from irony instances marked by hashtags arises from the potential variation in the linguistic behavior of posters. Since a hashtag itself often functions as a textual cue for irony, posters may consciously or unconsciously incorporate it when constructing their ironic utterances. In such cases, hashtags are effectively part of the ironic expression from a linguistic perspective. This, however, limits the variety of cues present in the corpus. For instance, such corpora often lack examples of “self-standing irony,” i.e., ironic expressions whose interpretation does not rely on explicit tags. Self-standing irony is crucial for text-based irony corpora for several reasons. For example, these cases more closely resemble spoken instances of irony in face-to-face communication. Additionally, as supported by our findings below, datasets created using tags are likely to conflate situational and verbal irony.

An alternative distant supervision technique that overcomes issues with the biased selection proposed by Shmueli et al. (2020). The method relies on reports of misunderstood irony in context, enabling the collection of a much wider variety of posts. Our work follows a procedure similar to theirs for collecting the potentially ironic posts. However, we do not simply assume that the distant supervision provides the correct messages, but we carefully annotate posts that are automatically selected.

corpus	source	annotation	ironic	non-ironic
Davidov et al. (2010)	English	Twitter, Amazon	tags	-
Ptáček et al. (2014)	Czech	Twitter	manual	325
Ptáček et al. (2014)	English	Twitter	tags	130 000
Oraby et al. (2016)	English	IAC	manual	3 260
Schifanella et al. (2016)	English	Twitter, Instagram, Tumblr	tags/manual	22 025
Peled and Reichart (2017)	English	Twitter	tags/manual	3 000
Karoui et al. (2017)	Arabic	Twitter	tags	1 733
Khodak et al. (2018)	English	Reddit	tags	1 340 000
Cignarella et al. (2018)	Italian	Twitter	manual	2 392
Ortega-Bueno et al. (2019)	Spanish	Twitter	manual	3 000
Castro et al. (2019)	English	video clips	manual	6 020
Oprea and Magdy (2020)	English	Twitter	manual	777
Shmueli et al. (2020)	English	Twitter	automatic	15 000
Ozturk et al. (2021)	Turkish	Twitter	manual	300
Misra and Arora (2023)	English	Onion/news	automatic	13 634
Qin et al. (2023)	English	Twitter	tags/manual	11 651
Sosea et al. (2024)	English	Twitter	manual	820
Ozturk et al. (2025)	Turkish	Satirical news	automatic	329

Table 1: A sample of corpora with irony annotations.

3. Source data

3.1. Misunderstood irony data

Our source data is collected from Twitter mainly through its streaming API.² Data was collected between March 2018 and February 2023 (until the free academic/research access to large amounts of Twitter data through Twitter API was revoked), by querying frequent words in Turkish, and storing the tweets that were also identified as Turkish by Twitter. We further filter the tweets for language using the Python langdetect library.³ This process resulted in approximately 3 billion tweets over 60 months, with an average of approximately 50 million tweets per month.

Using this large corpus of Turkish tweets, we follow a similar procedure to Shmueli et al. (2020) to identify the ironic Tweet candidates. We look for expressions of misunderstandings of irony such as *İroni yapıyor* ‘He/she is being ironic’ or *İroni yapıyordum* ‘I was being ironic’. The form of expressions that we observe for signaling misunderstood irony is much more varied than English – partially due to rich morphology, but also other expressions like *(Bu) ironi(ydi)* ‘this is/was irony’, *İroniyi anlamamışsın* ‘you missed the irony’ are also common. We use a set of regular expressions to catch these patterns.⁴ Unlike on English twitter, where irony/sarcasm is mostly signaled with expressions

²We refer to the platform with its old name, as our data comes from the time period before it was rebranded as X.

³<https://pypi.org/project/langdetect/>.

⁴The regular expressions and the scripts used for finding cue tweets can be found in our data repository at <https://github.com/coltekin/turkish-irony>.

userA: *Alarm sesimi o kadar seviyorum ki*
‘I love my alarm sound so much’

userB: *@userA nedir ki acaba :)*
‘what [sound] is it?’

userA: *@userB İroni yaptım userB*
‘I was being ironic, userB’

Figure 1: An example thread with misunderstood irony.

like ‘begin sarcastic’, the expressions listed above using the word *ironi* ‘irony’ is much more common in Turkish Twitter conversations. The word *sarkazm* ‘sarcasm’ (also *sarkastik* ‘sarcastic’) and their native form *iğneleme* can only be found in a few hundred instances within the whole corpus.

Once cues to misunderstood irony is found, we trace back in the conversation history to find the tweet with misunderstanding, and the tweet with probable irony. Figure 1 demonstrates the process with an example thread. We call the final tweet in the thread the *cue tweet*, and the tweet with misunderstanding the *oblivious tweet*. Furthermore, when possible we also collect *context tweets*, the tweets that were earlier in the conversation.

Besides the linguistic patterns for the cue tweets, we also check the user turns in the conversation to reduce the false positives. For example, for misunderstood irony self-clarified by the original author of the potentially ironic tweet (e.g., the expression ‘I was being ironic’ in Figure 1), we look for the closest earlier tweet to the cue tweet in the thread

that is written by the author of the cue tweet. Similarly, for a third person clarification (e.g., expression ‘He/she was being ironic’), we look for a tweet before the oblivious tweet that is posted by another author than the author of the cue and the oblivious tweet. In both types of cues, we also considered what post the oblivious tweet is responding to.

On a random sample of 200 cue tweets, [Shmueli et al. \(2020\)](#) reports over 97% accuracy for the precision of their cue patterns. In our experiments, likely related to higher variation of linguistic patterns in Turkish, we could not achieve similar precision without sacrificing recall. The final set of patterns we use to collect potential cue tweets has a precision of approximately 77% based on a manual inspection of 2 000 cue tweets. Since the data we have access to is much more limited, and our aim is not only relying on the cue tweets but annotating candidate irony tweets manually, we do not attempt to refine the patterns to increase the precision.

Another particular interest in the previous literature is also identifying *intended irony* and *perceived irony* ([Oprea and Magdy, 2019](#)). Commonly, the ironic social media posts that are tagged as ironic by their authors are considered intended irony, while the studies that are based on human annotations are considered as perceived irony. [Shmueli et al. \(2020\)](#) suggests that the present procedure identifies both, clarification done by the original author are considered as intended irony, and clarification posted by others are considered as perceived irony without further annotation. Although we mark the ironic instances similarly in our data set, we do not believe that the distinction is very clear and useful. First, these sets are not mutually exclusive. Most intended irony is also perceived as ironic at least by a part of the audience. And, perceived irony, in most cases, must be intended as irony. More interestingly, even though the current method identifies ironic expressions that are perceived as ironic by some of the audience, they are also perceived as non-irony by others. Second, since our interest eventually is to obtain reliable annotations based on annotators’ interpretation of expressions as ironic or not, we have a more reliable and naturalistic label for the cases of perceived irony. Finally, our cue patterns include expressions without clear indications of intended or perceived irony (e.g., *bu ironi* ‘this is/was irony’, or *ironiyi anlamamışsın* ‘you did not get the irony’). As a result, we provide a best-effort label for intended vs. perceived irony based on the cue patterns (also including an unknown label for the cases noted above). However, we do not further validate them, nor do we have claims of accuracy or usefulness of these labels.

Following the procedure above, we find 240 741

cue tweets. Our current data collection only relies on the self-collected Twitter stream described above. Since the data is necessarily incomplete, the full tweet history cannot be recovered for all cue tweets. As a result some of the parent tweets are not in our dataset, and the number of oblivious tweets and the number of potentially ironic tweets are smaller than the cue tweets. Tracing back the parents of these cue tweets result in 94 581 oblivious tweets and 15 411 tweets that are potentially ironic. Of the 15 411 potentially ironic tweets, 5 013 are root tweets. That is, they are not replies to another tweet. For the remaining 10 398, we extracted a single context tweet for 2 588 (43%) of them, two context tweets for 2 000 (33%) of them, three context tweets for 728 (12%) of them, and between four and 22 context tweets for the remaining 12%. The potentially ironic tweets come from 14 402 distinct authors, where the maximum number of tweets from one author is eight.

3.2. Irony-tagged data

In addition to potentially ironic tweets, we also collect self-reported ironic tweets by searching tweets tagged with irony markers (*#ironi* or *#ironik*) in the large tweet collection described in Section 3.1. Interestingly, we see rather few of these tags in the Turkish data, amounting to only 2366 instances of approximately 3 billion tweets. We also cannot find any other tags that similar to the tags that are used for marking irony and sarcasm in English tweets (e.g., *#sarcasm* and *#not* that are often used for building English irony/sarcasm datasets).

3.3. Random examples

Besides collecting potentially ironic tweets through cue tweets and hashtags, we also randomly sample some tweets for annotation. We sample these tweets from the tweets of the authors that are found in our potentially ironic tweet list to reduce the potential differences between ironic and non-ironic tweets because of other factors like linguistic style and thematic contents of the posts. We filter out those tweets that are in the potential irony sample, that contain irony tags, and that are shorter than five tokens excluding mentions, hash tags or URLs. The motivation behind this random set of tweets is not only providing more negative (non-irony) instances, but also being able to observe the rate of irony (as perceived by annotators) in a random sample.

4. Annotation process

We annotated two samples of 1 500 tweets from the data described in Section 3 above. For annotating one of the samples, we did not provide annotators

any context. In the second sample, we sampled only (potentially ironic) tweets with at least one parent in the thread, and the annotators were provided with all the preceding context available in the data. For the first sample, we randomly sampled an equal number of tweets from, potentially ironic tweets based on cues in the thread, the tweets that are explicitly tagged as ironic, and random tweet collection. For both samples, hashtags signaling irony were removed from the text before showing the tweet to the annotators. Since this may result in unnatural sentences and provides spurious cues to the annotators about the existence of the hashtag, we further filtered out the tweets where the hashtags are used as part of the sentence (e.g., *muhteşem bir #ironi örneği* ‘a great example of #irony’), selecting only instances where the hashtag is at the end of the tweet (possibly followed by other hashtags). This filtering left only about 100 hashtag marked tweets in the sample with context due to lack of tweets that matched the requirements. For the second sample, we complement the missing tagged irony tweets with the potentially ironic tweets identified through cue tweets. Hence, the second sample contains 100 tweets tagged as irony, 500 random tweets, and 900 potentially ironic tweets.

Our label set is essentially binary. We are only interested whether a tweet contains verbal irony or not. However, to encourage the annotators to make the distinction, and not to mark situational irony as verbal irony, we also included an additional label indicating situational irony. The annotators are asked to read the guidelines defining each label and providing examples before starting the annotation. They were encouraged to make a subjective decision based on their expectations when the tweet is ambiguous without further contextual information (e.g., political orientation or gender of the author of the tweet).

All 3000 tweets are annotated at least by three native speakers (the authors of the paper and a graduate student of linguistics), using a web-based annotation tool. The disagreements were revised by the annotators. However, we do not correct the annotations even if further information provided by the other annotators change the perspective of one of the annotators. Corrections are made only when an annotator realizes a clear mistake in their annotation. The inter annotator agreement is reported on the initial, non-adjudicated annotations. We also recruited additional voluntary annotators for a sample of 100 tweets that amounted to 10 annotations for each tweet, but our analysis here is focused on the annotations by the three annotators who annotated the whole corpus.

	without context		with context	
	2-way	3-way	2-way	3-way
cue	73.15	72.55	71.78	71.44
tagged	77.31	73.15	77.0	80.0
random	89.6	88.8	82.70	82.09
total	80.03	76.02	75.97	75.37
Fleiss' κ	56.88	55.63	58.04	57.61

Table 2: Agreement rates based on data source. The numbers indicate the posts where all three annotators agreed on. All numbers (including Fleiss' κ) are percentages. ‘cue’ indicates potentially ironic tweets identified by cue tweets, ‘tagged’ indicates tweets with hashtags, and ‘random’ indicates randomly selected tweets. The column labeled ‘3-way’ indicates the full label set including ‘situational irony’, and ‘2-way’ indicates the results after mapping ‘situational irony’ to ‘no irony’.

5. Analysis

5.1. Inter Annotator Agreement

For the first sample of 1500 tweets without context information, the three annotators fully agreed on 76.02% of the three-label decisions. Since we are only interested in verbal irony, and consider situational irony as non-irony, we also report the agreement on binary labels. The agreement on binary labels is 80.03%. Corresponding chance-corrected agreement scores (Fleiss' κ) are 55.63% and 56.88%, for three-label and binary setup, respectively.

For the second sample of 1500 tweets with additional conversational context, the rates of agreement are 75.97% for three-label choice, and 75.37% for the two-label choices. Corresponding chance-corrected agreement scores are 58.04 and 57.61

The detailed agreement rates based on the source of the target tweets are presented in Table 2. Not surprisingly, least agreement is observed on tweets collected through cues to misunderstood irony. Another expected difference is the increase in agreement rate when the ‘situational irony’ label is replaced with ‘no irony’. The effect is most visible with examples collected through hashtags. These tweets often describe situational irony rather than involving verbal irony themselves. This shows that hashtag-based datasets run the risk of conflating situational and verbal irony, reducing the possibility of distinguishing between them and obscuring key characteristics of verbal irony.

The drop of agreement in some cases when context is provided is contrary to our expectations. One would expect additional context to help an-

	without context			with context		
	I	N	S	I	N	S
cue	116	383	1	261	630	2
tag	124	330	46	44	53	3
random	25	474	1	53	442	2
	265	1187	48	358	1125	7

Table 3: Majority label distribution for annotations with and without context. The column label ‘I’ indicates verbal irony, ‘N’ indicates non-irony and ‘S’ indicates situational irony.

notators build the ‘common ground’ with the participants of the conversation, and be able to determine the ironic nature of the data more consistently. The differences in the makeup of the data (smaller number of hashtag-based tweets, larger number of misunderstood irony) may have been a possible explanation. However, the trend is visible with all sources (see Table 2). The pairwise agreement rates between annotators (not reported here) did not also reveal any indication of unexpected annotator behavior. One possible explanation is that the common-ground information necessary to correctly interpret irony in potentially ironic utterances may often be missing from the provided posts that precede the potentially ironic utterances. Such information may be inferred from broader non-linguistic contexts, such as the overall theme of the post or the personality traits of the posters. Annotators might mentally supply these broader contexts when immediately preceding linguistic contexts are not provided. Consequently, the presence of previous conversational turns (if they do not adequately support ironic interpretation) may have constrained the conceptualization of the broader context. In such cases, it may be possible that some annotators considered only the provided contexts, and some still considered the broader conceptualized context, leading to discrepancies in annotation decisions. We leave the further investigation of the cause of this discrepancy for future work.

5.2. Label distribution

Table 3 presents the majority-aggregated label distribution over both datasets. The first observation is that majority label assignment to verbal irony increases when more conversational context is present. Not only the tweets obtained through misunderstood irony, but also more tweets from other sources are annotated as ironic. In fact, the increase in the random sample is even more than the cue-based target tweets.

The rate of cue-based tweets that are inter-

preted as irony by our annotators is rather low. Even when more conversational context is available, the rate of irony among tweets that may have been marked as irony in a study that does not use manual annotation would be less than 30%. This may partially be due to the fact that our annotators are detached from the context of the conversation itself. Despite given conversational context, their knowledge of the participants in the conversation, the date/time of the discussions, and other aspects of common ground shared by the participants of the social media conversation is not available to our annotators. Hence, one may argue that it is not surprising that the number of posts that are interpreted as ironic by the annotators is low. However, closer inspection suggests that there are other reasons for the low rate of irony in these posts. One of the common reasons for the discrepancy is the loose use of term irony by social media users. It seems many social media users use ‘irony’ to express expressions that cannot easily be defined as irony, e.g., jokes or non-ironic hyperbole. Given that even experts in the field cannot fully agree on the definition of what irony is (Garmendia, 2018), it is not surprising that the social media users also use different definitions of the concept. Another common issue is the location of potentially ironic tweets in the thread. It is quite common that the cue tweet may be referring to a potentially ironic tweet much earlier than where the method expects to find an ironic tweet.

6. Conclusions

We presented a new Turkish social media corpus annotated for irony. We make use of distant supervision techniques to locate social media posts with a high probability of being ironic, and annotate them manually. The corpus created in this study is a useful resource for building automated systems for identifying irony, testing irony understanding capabilities of large language models, as well as for quantitative linguistic studies investigating irony identification and signals of ironic comments on social media. Our analysis also shows that relying only on distant supervision alone may result in suboptimal labels for irony. For future work, we plan to increase the size of our corpus as well as the number of annotations per tweet, and test the use of our corpus on irony detection on social media, as well as using it for testing the capabilities of large language models for irony understanding. The data is available at <https://github.com/coltekin/turkish-irony>.

Acknowledgements

This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) project number 538184825.

7. Bibliographical References

- Santiago Castro, Devamanyu Hazarika, Verónica Pérez-Rosas, Roger Zimmermann, Rada Mihalcea, and Soujanya Poria. 2019. [Towards multimodal sarcasm detection \(an _Obviously_ perfect paper\)](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4619–4629, Florence, Italy. Association for Computational Linguistics.
- Alessandra Teresa Cignarella, Simona Frenda, Valerio Basile, Cristina Bosco, Viviana Patti, Paolo Rosso, et al. 2018. Overview of the EVALITA 2018 task on irony detection in Italian tweets (IronITA). In *CEUR Workshop Proceedings*, volume 2263, pages 1–6. CEUR-WS.
- Dmitry Davidov, Oren Tsur, and Ari Rappoport. 2010. Semi-supervised recognition of sarcastic sentences in twitter and amazon. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, CoNLL '10, page 107–116, USA. Association for Computational Linguistics.
- J. Garmendia. 2018. *Irony*. Irony. Cambridge University Press.
- Jihen Karoui, Farah Banamara Zitoune, and Veronique Moriceau. 2017. SOUKHRIA: Towards an irony detection system for Arabic in social media. *Procedia Computer Science*, 117:161–168.
- Mikhail Khodak, Nikunj Saunshi, and Kiran Vodrahalli. 2018. [A large self-annotated corpus for sarcasm](#). In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).
- Yumin Kim, Heejae Suh, Mingi Kim, Dongyeon Won, and Hwanhee Lee. 2024. [KoCoSa: Korean context-aware sarcasm detection dataset](#). In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 9890–9904, Torino, Italia. ELRA and ICCL.
- Rishabh Misra and Prahal Arora. 2023. [Sarcasm detection using news headlines dataset](#). *AI Open*, 4:13–18.
- Silviu Oprea and Walid Magdy. 2019. [Exploring author context for detecting intended vs perceived sarcasm](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2854–2859, Florence, Italy. Association for Computational Linguistics.
- Silviu Oprea and Walid Magdy. 2020. [iSarcasm: A dataset of intended sarcasm](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1279–1289, Online. Association for Computational Linguistics.
- Shereen Oraby, Vrindavan Harrison, Lena Reed, Ernesto Hernandez, Ellen Riloff, and Marilyn Walker. 2016. [Creating and characterizing a diverse corpus of sarcasm in dialogue](#). In *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 31–41, Los Angeles. Association for Computational Linguistics.
- Reynier Ortega-Bueno, Francisco Rangel, D Hernández Farias, Paolo Rosso, Manuel Montes-y Gómez, and José E Medina Pagola. 2019. Overview of the task on irony detection in spanish variants. In *Proceedings of the Iberian languages evaluation forum (IberLEF 2019), co-located with 34th conference of the Spanish Society for natural language processing (SEPLN 2019)*. CEUR-WS. org, volume 2421, pages 229–256.
- Asli Umay Ozturk, Recep Firat Cekineli, and Pinar Karagoz. 2025. Make satire boring again: Reducing stylistic bias of satirical corpus by utilizing generative LLMs. In *18th Workshop on Building and Using Comparable Corpora*, page 19.
- Asli Umay Ozturk, Yesim Cemek, and Pinar Karagoz. 2021. IronyTR: Irony detection in Turkish informal texts. *International Journal of Intelligent Information Technologies (IJIT)*, 17(4):1–18.
- Lotem Peled and Roi Reichart. 2017. [Sarcasm SIGN: Interpreting sarcasm with sentiment based monolingual machine translation](#). In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1690–1700, Vancouver, Canada. Association for Computational Linguistics.
- Tomáš Ptáček, Ivan Habernal, and Jun Hong. 2014. [Sarcasm detection on Czech and English Twitter](#). In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 213–223, Dublin, Ireland. Dublin City University and Association for Computational Linguistics.

- Libo Qin, Shijue Huang, Qiguang Chen, Chenran Cai, Yudi Zhang, Bin Liang, Wanxiang Che, and Ruifeng Xu. 2023. [MMSD2.0: Towards a reliable multi-modal sarcasm detection system](#). In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 10834–10845, Toronto, Canada. Association for Computational Linguistics.
- Rossano Schifanella, Paloma De Juan, Joel Tetreault, and Liangliang Cao. 2016. Detecting sarcasm in multimodal social platforms. In *Proceedings of the 24th ACM international conference on Multimedia*, pages 1136–1145.
- Boaz Shmueli, Lun-Wei Ku, and Soumya Ray. 2020. [Reactive supervision: A new method for collecting sarcasm data](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2553–2559, Online. Association for Computational Linguistics.
- Tiberiu Sosea, Junyi Jessy Li, and Cornelia Caragea. 2024. [Sarcasm detection in a disaster context](#). In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 14313–14324, Torino, Italia. ELRA and ICCL.
- Lazar Đoković and Marko Robnik-Šikonja. 2024. [Sarcasm detection in a less-resourced language](#). In *Proceedings of Slovenian Conference on Artificial Intelligence 2024*. Jožef Stefan Institute.