Abdullayeva Iroda Murotovna, teacher
Uzbekistan State World of Languages University
ORCID: https://orcid.org/0009-0001-5592-3932

CORPUS-BASED ANALYSIS OF SATIRICAL TEXTS

Abstract. Satire is hard to analyze linguistically because it's meaning often lies in what is implied rather than stated. Corpus linguistics, with its focus on patterns across many texts, offers a systematic way to study how satirical effects are built, but counting words says little about irony unless the numbers are read in context. This article sets out how corpus methods can be applied to satirical texts. It discusses how to define and sample a satirical corpus, how to prepare and annotate the data, and how normalized frequency, keywords, collocation, semantic prosody, and concordance analysis can be combined. It argues that quantitative evidence is most useful when paired with close reading, and that differences between satirical and other texts should be interpreted with attention to genre, topic, and authors. A five-stage framework for future empirical research is proposed. No original corpus data are reported.

Keywords: *corpus linguistics, satire, irony, satirical texts, collocation, semantic prosody, concordance analysis, corpus design*

Annotatsiya. Satirani lingvistik jihatdan tahlil qilish qiyin, chunki uning ma'nosi ko'pincha to'g'ridan-to'g'ri aytilgan narsada emas, balki nazarda tutilgan narsada bo'ladi. Korpus lingvistikasi ko'plab matnlardagi qonuniyatlarga e'tibor qaratgani uchun satirik effektlar qanday yaratilishini tizimli o'rganish imkonini beradi, ammo so'zlarni sanash raqamlar kontekstda talqin qilinmasa, ironiya haqida deyarli hech narsa aytmaydi. Ushbu maqolada korpus usullarini satirik matnlarga qanday qo'llash mumkinligi bayon etiladi. Satirik korpusni qanday aniqlash va shakllantirish, ma'lumotlarni qanday tayyorlash va annotatsiyalash hamda normallashtirilgan chastota, kalit so'zlar, kollokatsiya, semantik prosodiya va konkordans tahlilini qanday uyg'unlashtirish mumkinligi muhokama qilinadi. Maqolada miqdoriy dalillar chuqur o'qish bilan birga qo'llanganda eng foydali bo'lishi, satirik va boshqa

matnlar o‘rtasidagi farqlarni esa janr, mavzu va mualliflarni hisobga olib talqin qilish kerakligi asoslanadi. Kelgusi empirik tadqiqotlar uchun besh bosqichli model taklif etiladi. Maqolada original korpus ma’lumotlari keltirilmaydi.

Kalit so‘zlar: *korpus lingvistikasi, satira, ironiya, satirik matnlar, kollokatsiya, semantik prosodiya, konkordans tahlili, korpus dizayni*

Аннотация. Сатиру трудно анализировать с лингвистической точки зрения, поскольку её смысл часто заключается не в том, что сказано прямо, а в том, что подразумевается. Корпусная лингвистика, ориентированная на закономерности, выявляемые на множестве текстов, позволяет систематически изучать, как создаются сатирические эффекты, но подсчёт слов мало что говорит об иронии, если числовые данные не интерпретируются в контексте. В статье рассматривается, как корпусные методы можно применять к сатирическим текстам. Обсуждаются определение и формирование сатирического корпуса, подготовка и разметка данных, а также сочетание нормализованной частотности, ключевых слов, коллокаций, семантической просодии и конкордансного анализа. Автор доказывает, что количественные данные наиболее полезны в сочетании с внимательным чтением, а различия между сатирическими и другими текстами следует интерпретировать с учётом жанра, тематики и авторов. Предлагается пятиэтапная модель для будущих эмпирических исследований. Оригинальные корпусные данные в статье не приводятся.

Ключевые слова: *корпусная лингвистика, сатира, ирония, сатирические тексты, коллокация, семантическая просодия, конкордансный анализ, дизайн корпуса*

1. INTRODUCTION

Satire is a kind of writing that uses humor, irony, and exaggeration to poke at people, institutions, and social habits. We meet it in newspaper columns, parody news websites, captioned cartoons, and comic sketches. In nearly all of them the text says one thing on the surface and leaves the reader to work out something sharper

underneath. That double layer is the reason satire has drawn the interest of scholars in linguistics, literary studies, and communication (Hutcheon, 1994; Simpson, 2003).

Corpus linguistics studies language through large, carefully collected sets of texts, called corpora, which are searched and analyzed with computer software. It is widely used to describe vocabulary, grammar, and discourse, and it has also been applied to humor (McEnery & Hardie, 2012; Partington, 2006). The big advantage is scale: a corpus lets a researcher look at thousands of sentences at once and see which patterns are typical of a kind of writing, instead of relying on a few memorable examples.

Satire is therefore an attractive but tricky subject for corpus work. Its meaning leans heavily on context and shared knowledge, so the numbers a corpus produces always need careful reading. This article looks at how corpus methods can be applied to satirical texts.

2. METHODS

2.1. Research design

This study is a conceptual, qualitative analysis. No corpus was compiled and no participants took part, so the article contains no original frequency data and needed no ethical approval. Its goal is to build, from existing scholarship, a framework that later empirical studies can put to the test.

2.2. Selection of sources

The analysis draws on thirteen works, chosen on purpose because they approach the problem from three sides. The first group deals with corpus methodology and the reading of frequency information (Sinclair, 1991; Baker, 2006; McEnery & Hardie, 2012; Gablasova et al., 2017). The second covers register variation and learner corpora (Biber, 1988; Biber & Conrad, 2009; Granger, 2009). The third focuses on satire, irony, and humor, including work on semantic prosody and automatic irony detection (Louw, 1993; Hutcheon, 1994; Attardo, 2001; Simpson, 2003; Partington, 2006; Reyes et al., 2013).

2.3. Analytical procedure

Each source was read with three questions in mind: what does it require of a corpus, what analytical tools does it offer, and what does it warn against? The answers were then compared across the three groups. Points that repeated or fitted together were grouped into stages, and this produced the framework presented in Section 3.4.

3. RESULTS

3.1. Requirements for corpus design

Four requirements stood out. The first is a clear definition of satire that is applied the same way throughout, for instance texts that label themselves as satire, or texts that readers and critics recognize as satirical (Simpson, 2003). The second is documented sampling. Sources, dates, authors, and topics should be recorded, because a corpus built from one outlet or one period mirrors the habits of that source and cannot stand for satire as a whole (McEnery & Hardie, 2012). The third is a comparison corpus matched for genre and topic. Satirical news is best set against ordinary news; setting it against fiction or conversation would mostly reveal differences of register (Biber, 1988; Biber & Conrad, 2009). The fourth is a written annotation scheme for devices such as hyperbole, understatement, mock praise, and rhetorical questions, with definitions and, where possible, more than one annotator.

3.2. Analytical techniques

Four techniques worked well together. The first, **normalized frequency**, shows how often a feature occurs and makes corpora of different sizes comparable (McEnery & Hardie, 2012). A common measure is frequency per 1,000 words:

Suppose, hypothetically, that a satirical corpus of 80,000 words contains 240 occurrences of an intensifier. Its normalized frequency would be 3 per 1,000 words. This example only illustrates the calculation and is not an empirical finding. Normalization does not show how evenly a feature is spread, so the number of texts and authors behind a pattern should be reported as well (Gablasova et al., 2017).

The second technique, keyword analysis, picks out words that are unusually frequent compared with a reference corpus. What it finds depends directly on that reference: a general corpus produces many keywords that simply reflect topic and genre, while a matched news corpus is more likely to point to satirical purpose (Gablasova et al., 2017). The third, collocation and semantic prosody, ties a pragmatic effect to a pattern that can be searched. Words tend to keep certain company (Sinclair, 1991), and when a word turns up in a setting that clashes with its usual evaluative flavor, the clash can signal irony (Louw, 1993). The fourth, concordance reading, shows every occurrence in its context. Picture an invented parody headline that congratulates a city council on repairing a road "in record time" after ten years. A frequency count would see only ordinary words, whereas the concordance line shows the mockery.

3.3. Interpretive safeguards

Three safeguards were identified. First, alternative explanations deserve a fair hearing. A satirical corpus may contain more intensifiers than a news corpus because of satirical exaggeration, but also because of a more opinionated register, different topics, or a few prolific authors. Second, the spread of a feature across texts and authors should be checked before it is called characteristic of satire. Third, a single feature is a hint, not proof. Research on automatic irony detection combined several kinds of features and still found that no single one was decisive (Reyes et al., 2013), which supports this caution.

3.4. Proposed framework

The requirements, techniques, and safeguards were organized into five stages, summarized in Table 1.

Table 1. Five-stage framework for the corpus-based analysis of satirical texts

Stage	Key decision	Technique	Safeguard
1. Define	What counts as satire	Explicit selection criterion	Apply the criterion consistently and report it
2. Sample	Sources, dates, authors, topics	Documented sampling; matched comparison corpus	Do not generalize beyond the sampled sources

Stage	Key decision	Technique	Safeguard
3. Prepare	Cleaning and annotation rules	Tagging; written annotation scheme	Report annotator agreement
4. Analyze	Which features to examine	Normalized frequency, keywords, collocation, concordance	Report distribution across texts and authors
5. Interpret	What the pattern means	Weigh alternative explanations	Treat single features as suggestive

4. DISCUSSION

The results suggest that the real difficulty of corpus research on satire is interpretive, not technical. Corpus tools work on form, while satire works on meaning in context, and the same word can be ironic in one sentence and sincere in the next (Hutcheon, 1994). For this reason, the framework does not ask the corpus to detect irony. It uses the corpus as a disciplined way to find and check patterns, in line with the wider view that corpus evidence guides interpretation and does not replace it (Baker, 2006; McEnery & Hardie, 2012). The stress on matched comparison corpora comes from register research, which shows that genre and situation shape language as strongly as communicative purpose does (Biber & Conrad, 2009).

The framework may also help in language teaching. Satire is hard for readers who do not share the cultural background it assumes, and that includes learners of English as a foreign language. Findings about recurring evaluative vocabulary, exaggerated quantifiers, and unexpected word combinations could become the basis of reading activities that train learners to notice signs of irony. As Granger (2009) argues for learner corpora more generally, such findings should be used with care and weighed against what teachers know about their own learners. A list of "satirical features" should never be handed out as a fixed formula, because the effect of satire always depends on context.

Some limitations are worth stating plainly. The framework is conceptual and has not yet been tested on data. The sources were chosen on purpose and do not amount to a systematic review. Satirical meaning rests on shared knowledge that a

corpus rarely records, and the annotation of devices involves interpretation. Future studies could apply the framework to well-documented corpora of satire from several sources and languages, compare satirical texts with matched non-satirical ones, and combine corpus work with reader-response data, for example by asking readers to mark where they sense irony and then examining the language of those passages.

5. CONCLUSION

This article has proposed a five-stage framework for the corpus-based analysis of satirical texts, built on four requirements for corpus design, four complementary techniques, and three interpretive safeguards. Its main point is simple: numbers and close reading work best together. Used that way, a corpus can show how satirical effects are assembled from ordinary linguistic resources, and where those resources are bent to say something sharper than the surface suggests. The next step is to test the framework on documented corpora.

REFERENCES

Attardo, S. (2001). *Humorous texts: A semantic and pragmatic analysis*. Mouton de Gruyter.

Baker, P. (2006). *Using corpora in discourse analysis*. Continuum.

Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.

Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge University Press.

Gablasova, D., Brezina, V., & McEnery, T. (2017). Exploring learner language through corpora: Comparing and interpreting corpus frequency information. *Language Learning, 67*(S1), 130–154. https://doi.org/10.1111/lang.12226

Granger, S. (2009). The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation. In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 13–32). John Benjamins.

Hutcheon, L. (1994). *Irony's edge: The theory and politics of irony*. Routledge.

Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 157–176). John Benjamins.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

Partington, A. (2006). *The linguistics of laughter: A corpus-assisted study of laughter-talk*. Routledge.

Reyes, A., Rosso, P., & Veale, T. (2013). A multidimensional approach for detecting irony in Twitter. *Language Resources and Evaluation, 47*(1), 239–268.

Simpson, P. (2003). *On the discourse of satire: Towards a stylistic model of satirical humour*. John Benjamins.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.