User Comments across Platforms and Journalistic Genres: A Computational Analysis

Anat Ben-David   anatbd@openu.ac.il   ✳   Oren Soffer   orenso@openu.ac.il

THE OPEN UNIVERSITY OF ISRAEL

Omilab

Introduction

Using computational methods, this study compares user comments to the same news articles across three platforms: the commenting section on a newspaper's website, the comments posted by Facebook users on the newspaper's website through Facebook Plugin, and comments on the newspaper's Facebook page. Since each commenting channel is embedded in a different media environment, a comparative cross-platform approach allows us to gain understanding about the interaction between news content, user comments, and online platforms. While previous cross-platform studies of comments on news mainly focused on the anonymity of comments on news websites compared to the highly identified environment of Facebook comments (cf. Santana, 2014), the use of computational tools allows us to further characterize differences in commenting behavior across platforms.

Method and Tools

Our analysis focuses on the popular Israeli website Ynet. We compare comments to articles from the (pre-moderated) commenting section and in the Facebook Plugin commenting box on Ynet's website, and from the comments on Ynet's Facebook page. Overall, we analyzed 60 articles and 17,437 comments from July to December 2015. We built server-side custom tools to extract data from Facebook's API and from Ynet:

- **Facebook Comments Scraper:** The researcher can search for a public Facebook page and select a specific time range. The tool then extracts all posts. The researcher selects posts for comment scraping. The tool then scrapes all comments. (See Figure 1.)
- **Ynet Comments Scraper:** The researcher types a URL of a news article published on Ynet, and subsequently generates a script that recalls the comments at the bottom of the article, as well as from Facebook Plugin. (See Figure 2.)

- Both tools output the extracted comments in tabulated files and automatically compute each comment's length.
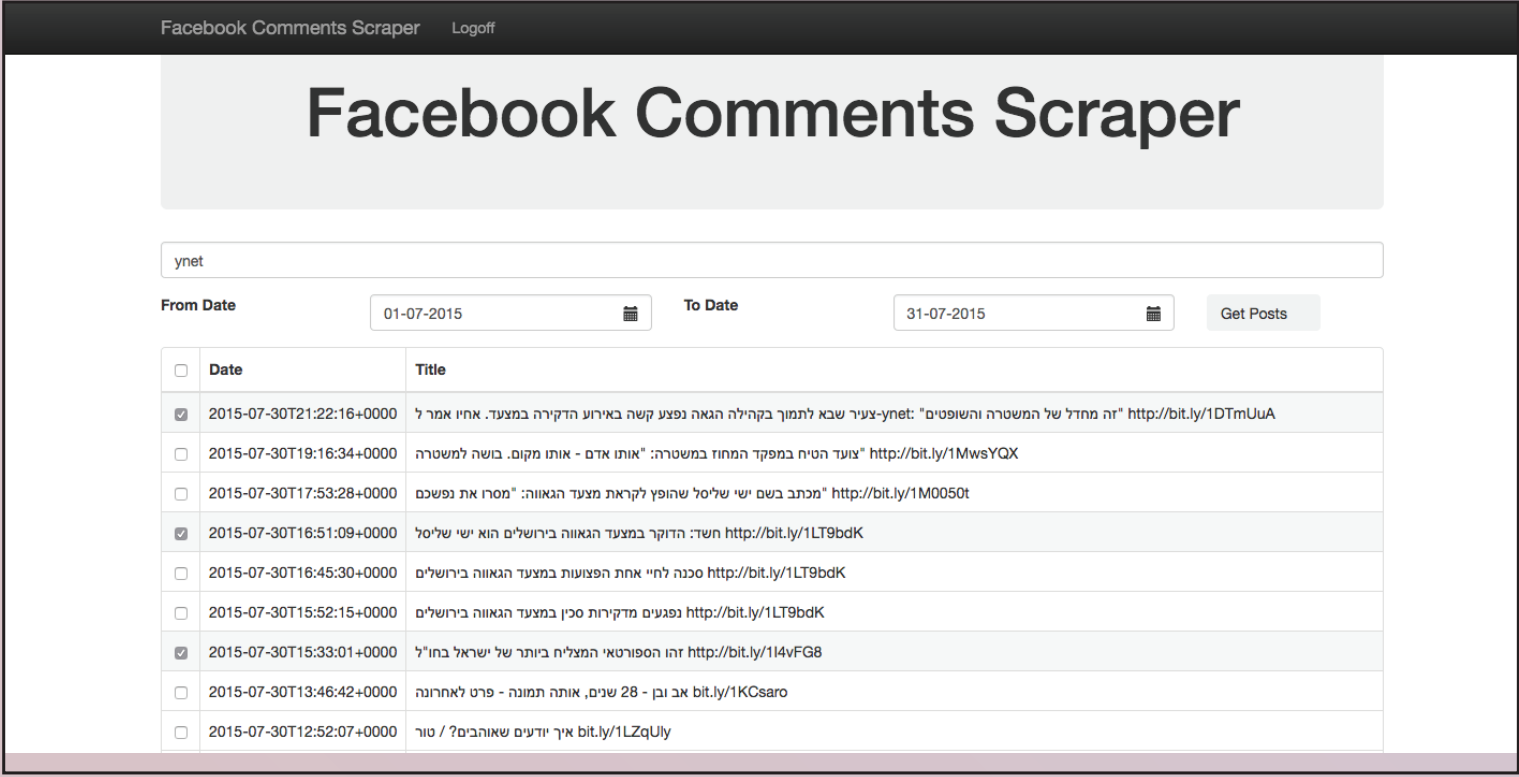- Per article, the tools compute the total and average commenting count and length.

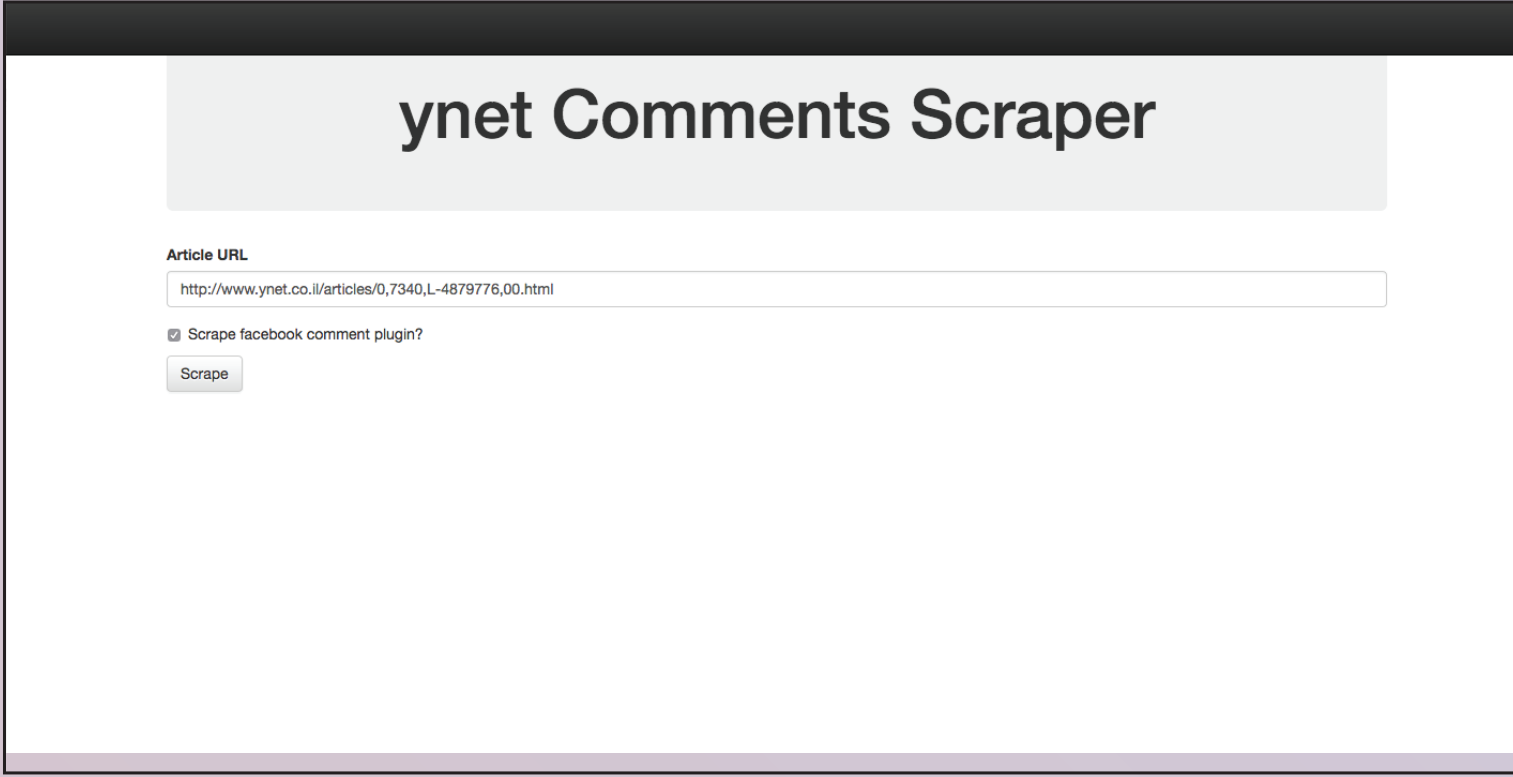Figure 1. *The User Interface of the Facebook Comments Scraper*

Figure 2. *The User Interface of the Ynet Comments Scraper*

To further analyze the content of the extracted comments, we grouped the comments' text per platform and per news type, following the theoretical distinction between hard and soft news (Boczkowski, 2009). After addressing issues of stemming and morphological disambiguation of Hebrew texts, we used Latent Dirichlet Allocation (LDA) topic modeling to extract topics in the comments to all news in each of the studied platforms, per news type. Interactive topic models were generated using LDAvis (Sievert and Shirley, 2014).

Results: Our findings show statistically significant differences in commenting patterns across platforms and journalistic genres:

● Comment Count

Across platforms, the number of comments to the same news items on Facebook is almost double than the number of comments on Ynet, and almost ten times higher than the number of comments posted through Facebook Plugin.

Across Genres, there are more comments to hard news than soft news on Ynet and Facebook Plugin (T=2.418, F=5.719, P<0.05). (See Figure 3.)

Across platforms and Genres, there is a similar proportion of comments to hard news on Ynet and through Facebook Plugin (about 78.1% of the total comments), compared to Facebook, where comments to hard news items make 60.4% of the total comments, (F=20.195, P<0.01).
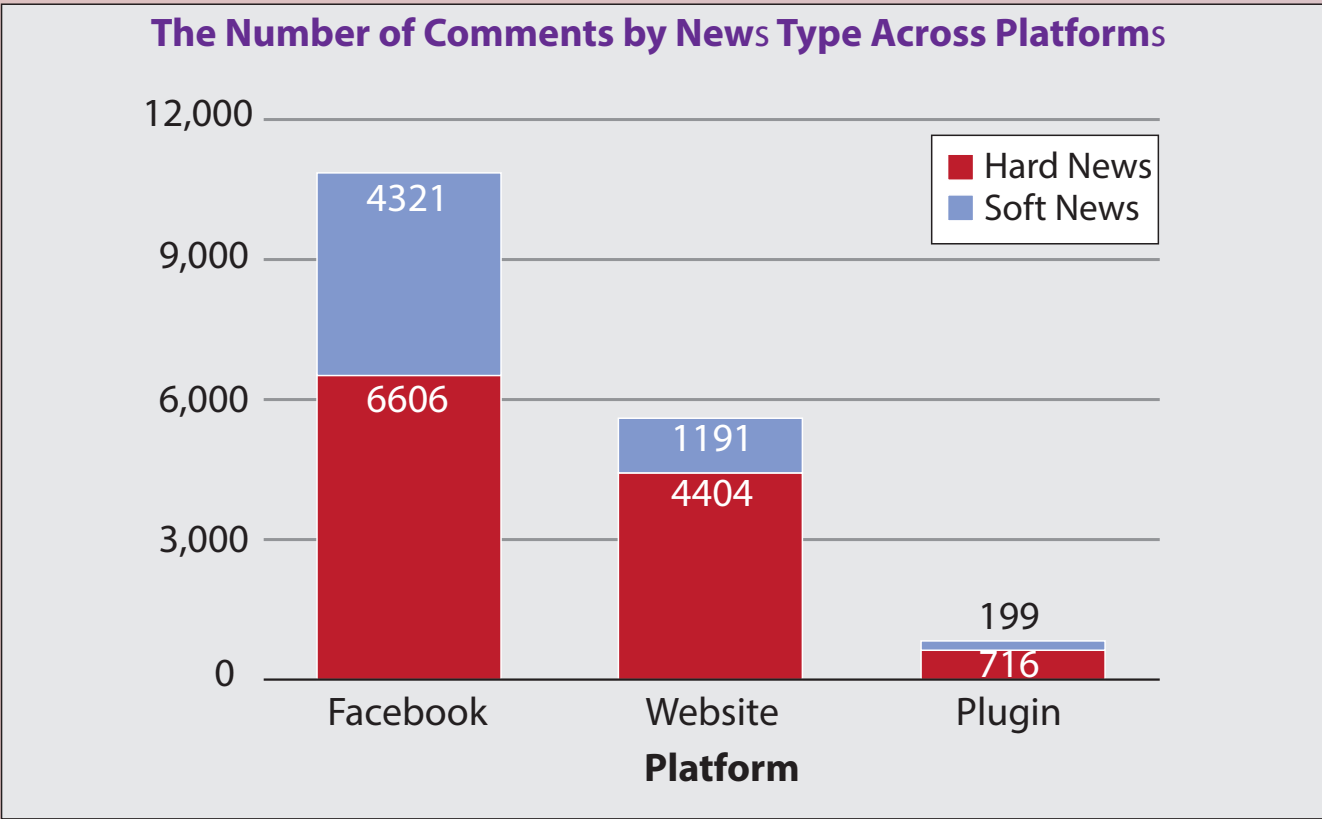

Figure 3. *The Total Number of Comments Aggregated by News Type and Platform*

● Comment Length

Across platforms, on average, the comments posted on Facebook are shorter compared to the comments posted Ynet. However, comments posted via Facebook Plugin are strikingly longer than the two other platforms, indicating a performative and deliberative behavior (F=4.635, P<0.05).

Across Genres, comments to hard news items are nearly two-thirds longer than comments to soft news items (T=3.163, F=5.182, P<0.05). (See Figure 4.)
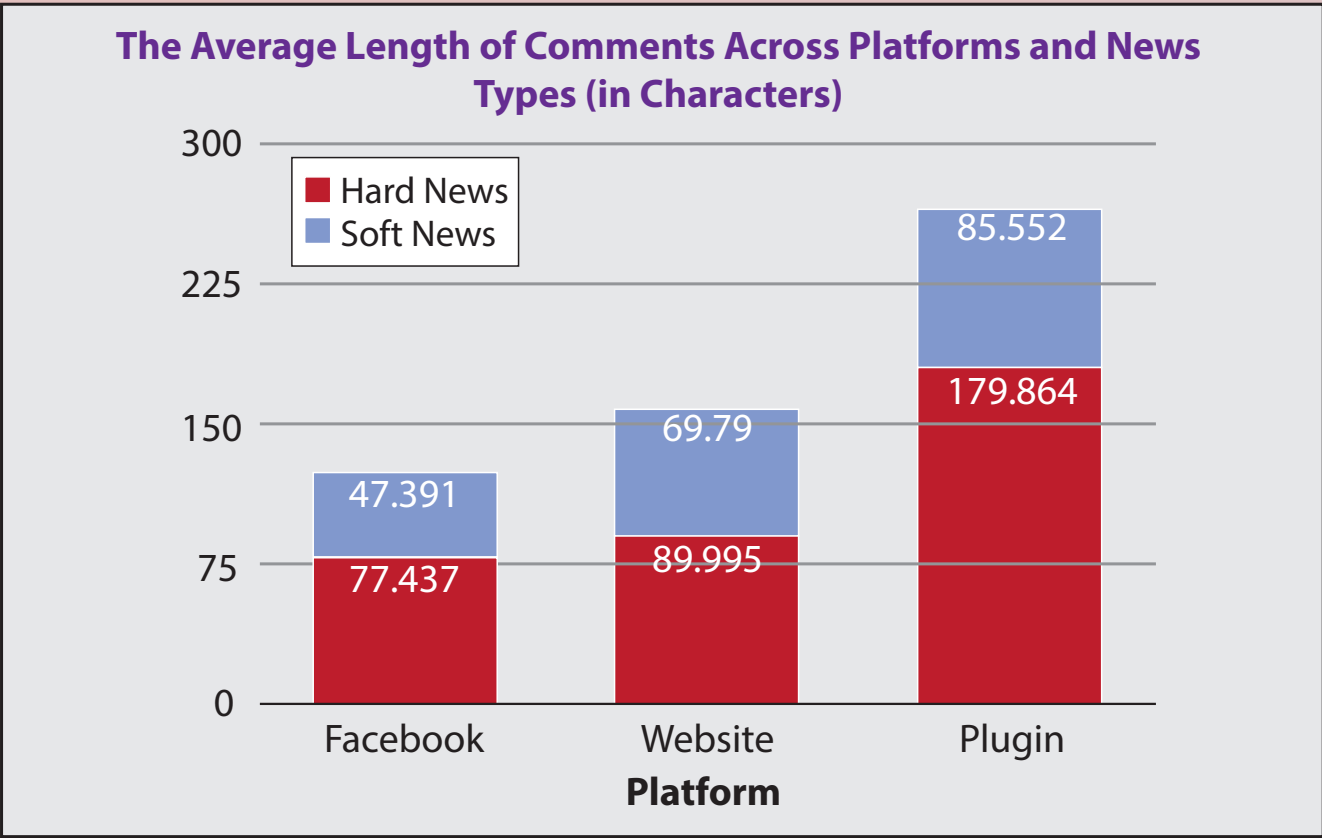

Figure 4. *The Average Length of Comments (in Number of Characters), Aggregated by News Genre and Platform*

● The Life Cycle of Comments on Facebook across Journalistic Genres

The mean commenting time to hard news is almost 5 hours after an article is posted (295.89 minutes), compared to a mean of nearly 7.5 hours to soft news (442.67 minutes), (F=86.544, T=-8.131, P<0.01).

Comments to hard news tend to peak in the first 5 minutes after an article is posted and then gradually decline, whereas comments to soft news are characterized by several gradually declining peaks, with intervals of less commenting activity between them. After the first half hour, the commenting rate of hard news sharply decreases, compared to a moderate decrease in the commenting rate of soft news (see Figure 5).
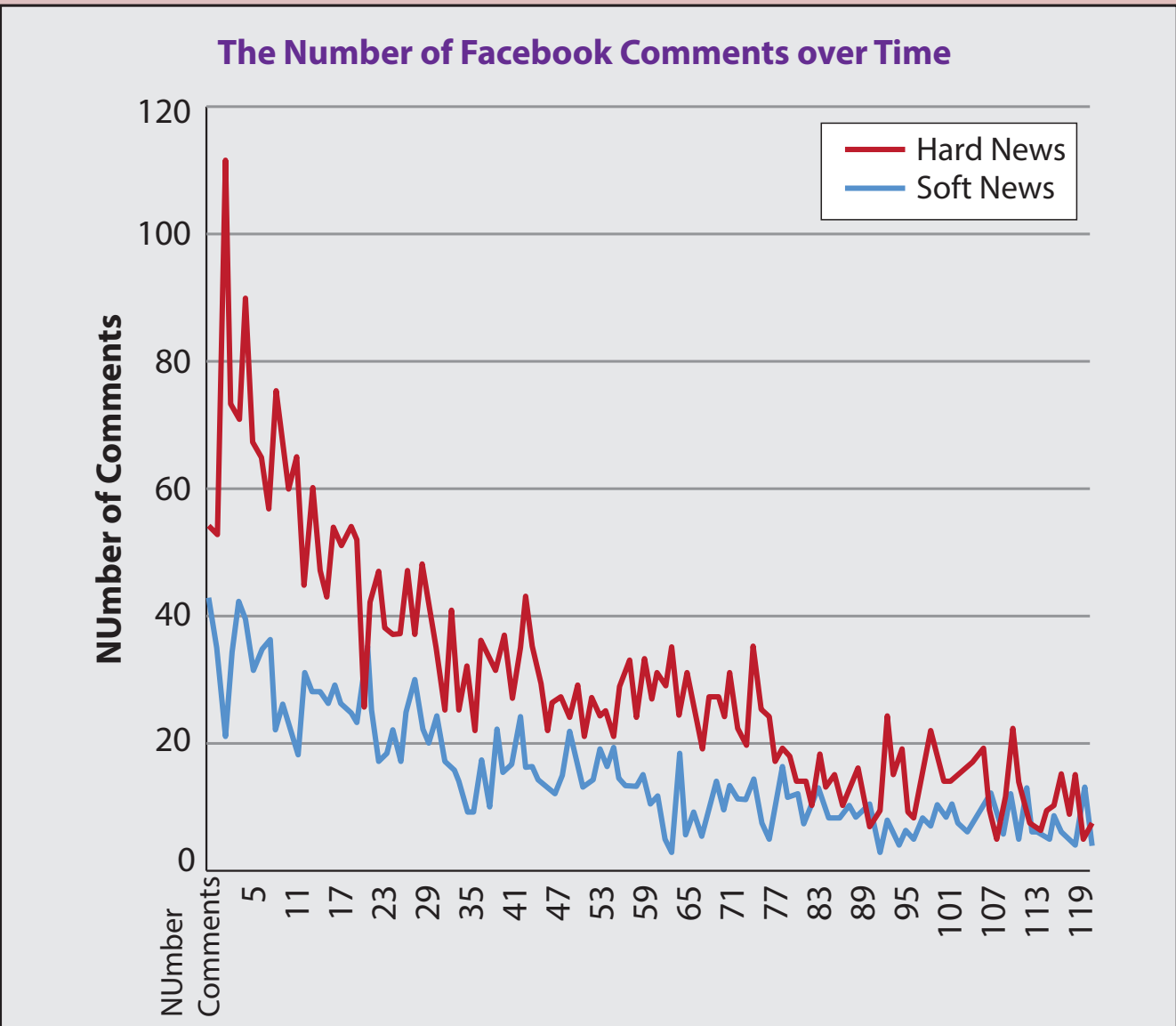

Figure 5. *The Number of Comments per minute to hard and soft News Articles on Facebook – A view of the first 120 minutes*

● Topic Modeling

There are distinct commenting topics to the same news articles across platforms and genres. Facebook comments to hard news contain negative emotional elements, expressing sadness, anger, and grief related to terrorism and security issues (see example in Figure 6).
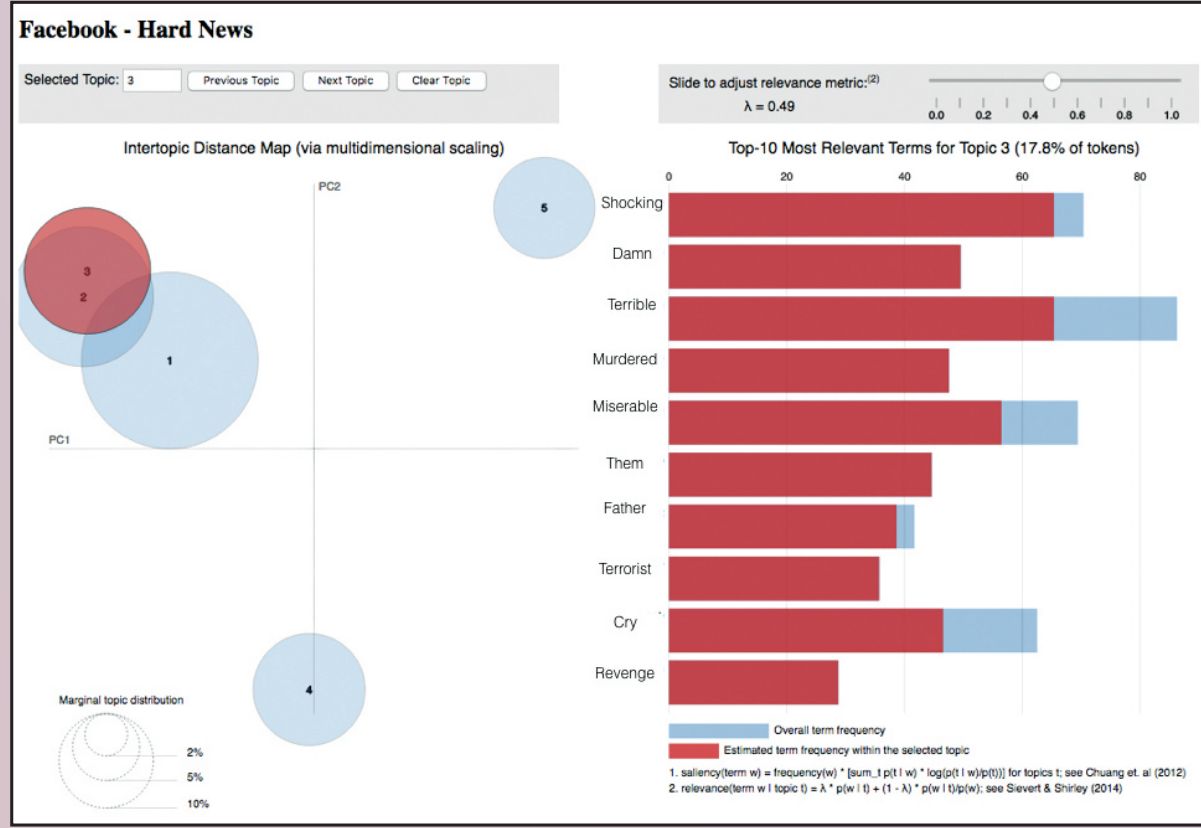

Figure 6. *Terms comprising the third topic of comments to hard news on Facebook. Topics 1–4 deal with security and terrorism. Topic 4 deals with economical issues*

Topics in comments to hard news posted on Ynet relate to security and terrorism, but they also deal with other domestic issues relating to the economy, the government, and the Iranian nuclear weapon program (see example in Figure 7).
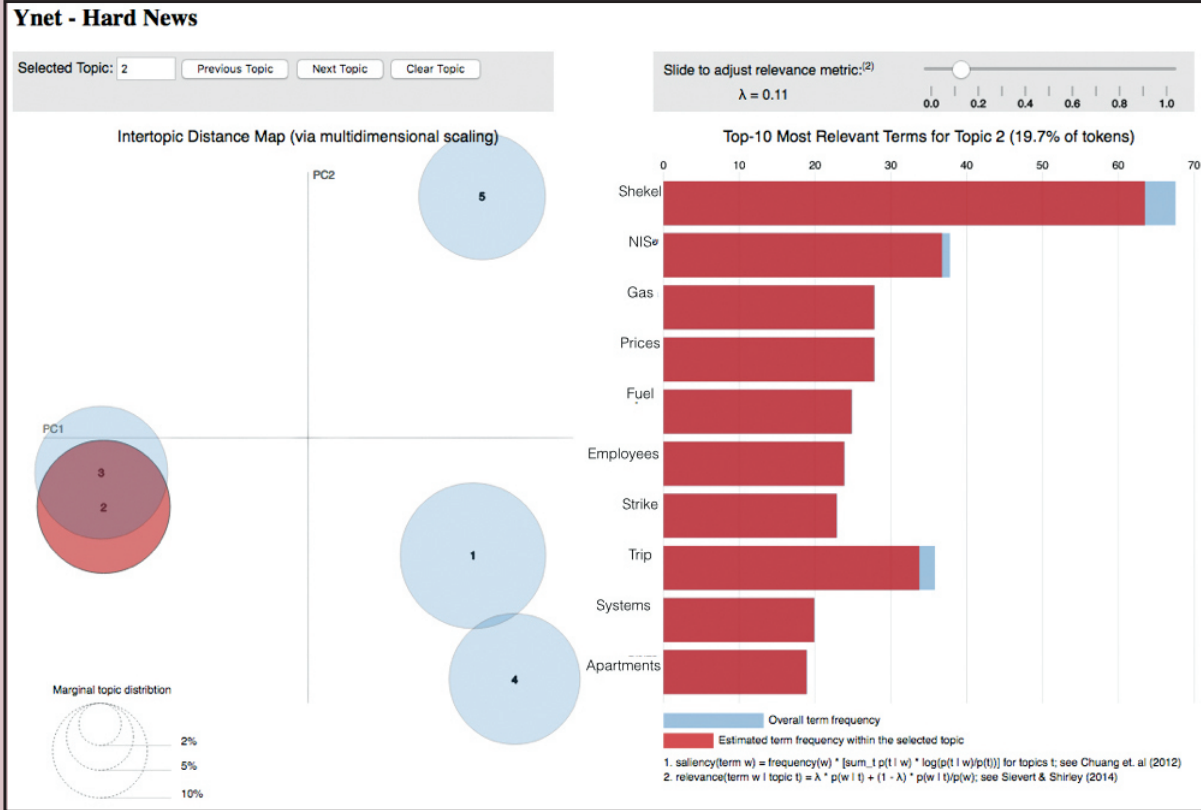

Figure 7. *Terms comprising the second topic of comments to hard news on Ynet. opics 1, 4 and 5 deal with security and terrorism. Topics 2 and 3 deal with domestic economic and social issues*

Interestingly, topics extracted from Facebook Plugin comments to hard news items resonate more items on international affairs, rather than domestic issues (see example in Figure 8).
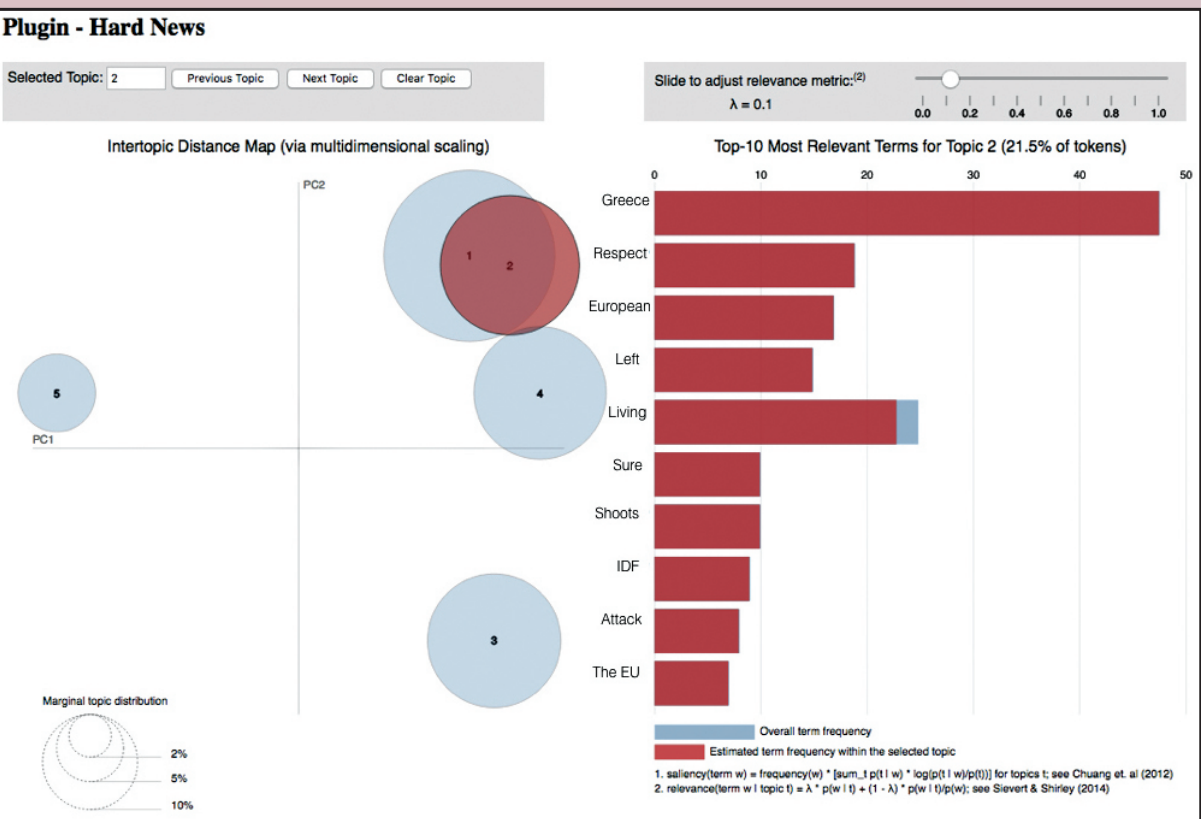

Figure 8. *Terms comprising the second topic of comments to hard news posted via Facebook Plugin. Topics 1–2 deal with international affairs. Topics 2 and 3 deal with domestic economic and social issues*

Overall, the topics extracted from all comments exhibit a strong element of national identity. A recurring word in topics across platforms and genres is the pronoun "we", or "us", which can be seen as a banal marker of nationalism (Billig, 1995). The "us" in this case relates to the Jewish-Israeli national identity, as opposed to the interpolated "them"—the Arabs, Palestinians, Iran, the UN, the US, the EU, or the world in general.

Discussion

Cross Genre Analysis: Hard and soft news—terms originally related to news production and editorial decisions—were found relevant in characterizing online user comments. Our findings indicate that hard news trigger more comments, and the cycle of comments to hard news is distinct from the cycle of comments to soft news. This confirms previous studies that have found that news stories on controversial political/social issues receive the highest number of comments (Boczkowski & Mitchelstein, 2013, p. 135).

Cross platform analysis: While our analysis is limited to commenting features that are available, measurable and shared by the three studied platforms, the comparison of comments to the same content across platforms allows us to characterize platforms as contextual environments that shape commenting cultures. Our different analyses show the prominent place reserved to social media in people's engagement with news.

The differences found in the number of comments between the news website and its Facebook page not only relate to quantifying readership, but may also be an outcome of comment pre-moderation on the news website. Thus, the comment moderation process in popular news sites such as Ynet—which receives vast numbers of comments—might result in a high rate of comment rejection. As a result, the public opinion climate, which is reflected through reading the published comments in the comments section, is not identical to the one that would have been reflected through reading all posted comments.

Against the high number of comments both on Facebook and on Ynet, the paucity of comments posted through Facebook Comment Plugin can be explained by the hybridity of this social media feature.

Conclusions

User comments to the same journalistic contents vary greatly in form and content across platforms. User comments to news articles are affected by each platform's cultural practices and technological affordances, which, in return, shape the public discussion of news. Focusing on one platform alone would miss out important contexts, emphasis, dynamics and interactions with the same content that take place simultaneously on a different platform. Thus, to understand user comments' standpoint to specific news content, we must take into account multiple heterogeneous contexts across platforms.

Contribution to the Hebrew speaking research community:

At the Open Media and Information Lab of the Open University of Israel, we build computational tools that are specifically designed to support challenges in the Hebrew language. Like other Semitic languages, Hebrew has unique characters, morphological structure, word order, and writing direction. In many cases, the application of existing tools and software to analyze Hebrew texts extracted from the Web renders illegible text, and textual or semantic analysis tools often cannot be applied. While there are advances in the computational analysis of Hebrew texts in the fields of Computational Linguistics and Digital Humanities, there is still an acute shortage of available methods and tools for Social Scientists interested in analyzing data in Hebrew, especially in the context of communication and media research. The tools built for this research will be made available to Israeli social scientists who wish to scrape and analyze data in Hebrew from Facebook / Ynet.

Acknowledgements

Works Cited

Billig, M. (1995). Banal Nationalism. London: Sage.

Boczkowski, P. J. (2009). Rethinking Hard and Soft News Production: From Common Ground to Divergent Paths. Journal of Communication 59(1): 98–116.

Boczkowski, P. J., & Mitchelstein, E. (2013). The news gap: When the information preferences of the media and the public diverge. Cambridge: MIT Press.

Santana, A. D. (2014). Virtuous or vitriolic: The effect of anonymity on civility in online newspaper reader comment boards. Journalism Practice, 8(1), 18–33.

Sievert, C., & Shirley, K. E. (2014, June). LDAvis: A method for visualizing and interpreting topics. In Proceedings of the workshop on interactive language learning, visualization, and interfaces (pp. 63–70).