

## Language investigation: Language and social media

The growth of social media has changed both how we communicate with each other including the percentage of our communication that takes place online. It's changed who we interact with, in many cases creating broader social networks. Many of us now connect on a regular basis with a much larger network of people than we did in the past. We have access to the lives, interests, thoughts and feelings of a wide range of people. They may be our best friends, family, someone we met once at some event somewhere or just someone with a common interest such as social justice, posting photos of your outfit of the day, or finding humour in cat memes and doge.

Social media platforms have evolved substantially from the days of Bebo and MySpace through Facebook and Twitter and now Snapchat and Instagram. By the time you're reading this, something new will no doubt be gaining popularity and taking its own share of the social media market. Think about the following questions in relation to the changing landscape of social media:

1. In what ways has social media changed in the last 5-10 years?
2. What are some key differences between:
  - a. Bebo and MySpace (if you've even heard of them!);
  - b. Facebook and Twitter;
  - c. Snapchat and Instagram;
  - d. any other platforms you use?
3. What does content look like on each platform?
4. Have older platforms copied some of the innovations of the new? Try to think of examples.

So how do people use these platforms? Are they all used in similar ways? Does it depend on who's using them? How do the types of content, the platform on which content is posted, and the characteristics of the content poster affect both who its audience is and how they interact with it? How do platform, genre and social factors affect language use (e.g. standard versus non-standard grammar)? These are just a few questions linguists interested in exploring social media might ask.

### Finding a focus and gathering data

In order to investigate some of the questions above, you need to gather some data. In addition to Facebook, Twitter, Instagram and Snapchat, you might also want to consider looking at YouTube videos. While the relationships and interaction between content poster and audience are often different on YouTube, it still has strong social media aspects. It's also a common focus of studies in linguistics, both the videos themselves and their comments sections.

Here we will look at how one or multiple factors about the content and the poster of the content affect both the content's form and its audience reception.

### A note on ethics

Be mindful of ethical issues when collecting data from social media. In general a minimum requirement is that data you collect for analysis has been shared publicly. Posts from your friends on Facebook and sometimes Twitter and Instagram may have only be shared with approved connections, for instance. Reproducing such posts in studies and thereby disregarding their intended audience is unethical (unless you gain the poster's permission). You could analyse your own social media posts (regardless of privacy settings), however if you wish to analyse comments the same ethical consideration applies. Even if data is public, it's also good practice to anonymise data where possible.

### Variables

It's up to you whether you want to do any quantitative analysis here. Regardless, it can be useful to think about the focus of your investigation in terms of *independent* and *dependent variables*. This can help establish clarity on what it is that you actually want to investigate. Say you want to look at the length of video or photo descriptions (or captions) on Facebook vs Instagram. Another way of phrasing this is "the effect of the platform on caption length". The thing that you're exploring the effect of is the independent variable (platform), while the thing it has an effect on is the dependent variable (caption length).

Unless you already have a clear idea of what you want to research (perhaps based on previous observation of an interesting pattern in social media use), you can start with either independent variable(s) or dependent variable(s) and then work out what might be of interest from there. Below are some suggested independent and dependent variables. Feel free to come up with your own.

Suggested independent variables (things that have an effect):

- **Genre.** In a study of the Oscar Pistorius trial and YouTube's role in journalism, Ruth Page broke down the genre of videos into four categories: mainstream news reports; news commentary; vlogger responses; parody songs. Think about the different types of genres content providers might adopt in covering the same topic.
- **Platform:** e.g. YouTube vs Instagram vs Snapchat
- **Social factors about the poster:** e.g. age; disability; gender; race; sexuality. You might also consider accent, which can convey information about a speakers' regional and social class background.

Suggested dependent variables (things that are affected):

- **Language:** e.g. standard vs non-standard grammar and orthography (e.g. "Are you okay?" or "u ok"), including whether regional features of spoken language are reflected in written language (e.g. "Scottish Twitter").
- **Audience reception:** how much do people engage? What types of engagements do they make: Facebook reacts or likes/loves/heart reacts on other platforms; text-based, image or GIF comments; shares? What is the content of their response: positive/negative; supportive/challenging; humorous/serious?
- **Content type:** video, photo, text, link, something else or a mixture of these.
- **Content purpose:** e.g. information, entertainment, creative expression or self-promotion. There may be some overlap between this and genre.

Some factors could be both independent and dependent variables, such as content type. You might also want to look at the effect of genre or platform on the topic, or you might want to look at the effect of the topic on audience reception/types of audience engagement.

### Selecting the data

Once you've picked one or two independent variables, you need to select data to analyse. You need to make sure you have enough data to make comparisons within each independent variable, and potentially – for a more nuanced “interactional” analysis – within each combination of independent variable. Say you want to look at gender and genre within a single platform (e.g. Facebook). Create a cross-table where each category of gender you wish to explore is a row heading and each category of genre is a column heading. You shouldn't aim to create an exhaustive list of genders or genres (that would be impossible): just focus on a few categories within each independent variable. Using broad umbrella categories (like “non-binary” or “opinion” below) can help you boost the numbers but be aware that you may miss meaningful differences between smaller, more specific categories (e.g. opinions on politics vs opinions on reality television). Your table might look something like this:

	Opinion	Meme	Event videos/photos
Female			
Male			
Non-binary			

Your aim now is to gather enough data to have – ideally – five examples within each cell of your table. Note that if you don't want to look at the interaction of independent variables (how they work together) and instead just look at each in isolation, you can create two tables, one with gender column headings and one with genre column headings. This might mean you need less data overall to get five examples in each cell, since most examples will count for a cell in each table. Depending on where you gather the data from, you may also have to drop some categories, e.g. you might not find (enough) openly non-binary people in every social domain. Be aware that for many social characteristics – e.g. class, disability, gender, race and sexuality – it's not always possible to tell how to categorise someone based on their name, how they look or how they sound. Self-declaration is the best evidence, but in the absence of this you may have to make some assumptions. Make sure to be up front about any assumptions you do make!

### Analysing the data

Now you have your data, it's time to start taking notes and making observations. In some cases this might be a simple categorisation of data. If your dependent variables are content type and content purpose, you could come up with a few categories of each, create a table and categorise each example within the dataset. Audience reception could also be quite categorical (see above), or you might want to take more detailed qualitative notes on each example and perhaps consider both text and non-text engagement and if/how they interact.

Once you've categorised all your data and taken notes where appropriate, start to look for patterns. How do the independent variables affect what you've found when recording information about the dependent variables?

**What do you think?**

Are there meaningful patterns in your data? Might other independent variables be more or less important?

What else might be of interest? You could look at direct messaging across platforms (with the consent of all participants) or perhaps more text-based platforms like Reddit (i.e. forums). Is there a difference between platforms like Reddit, platforms like YouTube and more personal network-based platforms like Facebook, Twitter, Instagram and Snapchat?