How a Computational Lexical ANALYSIS CAN HELP A BRAND

Contributed by

ASHLI SHERMAN
VP, Client Services,
Ogilvy CommonHealth Behavioral Insights, part of Ogilvy CommonHealth Worldwide

CHRIS CULLMANN
Senior VP, Engagement Strategy,
Ogilvy CommonHealth

Data continues to be a focal point for marketers in 2017. It informs our decisions, allows for increasingly accurate insights, and helps measure success. More so now than at any other time in the history of advertising, "data" is driving the strategies and tactics for creators. It is becoming very easy to get lost in the technological drivers and the manifestation of "data" when creating a campaign or messaging. Equally as challenging is how to effectively discuss leveraging research, observation, insights, and measurement for healthcare communications.

To take advantage of data in our programs, we need to consider that "data" is not new. We have been using insights and research in our marketing for generations. The change is coming from the volume and speed with which the data is being made available to us. Let's consider when developing a marketing campaign, new and innovative content, or any program: have you ever wondered things like, "What do I need to say to emotionally engage my target audience?" or, "What is the best way to position a new brand expected to come to market?" or, "What is the best word or phrase to be using — and what should I avoid?" The choices made are typically based on research, business decisions, and inference. Here is an excellent example of putting data to work for the marketer. Beginning with specific language, let's consider how we would build a program that was based on assessing the lexicon to make the campaign a whole lot stronger — and in the end, improving the dialogue around the brand between healthcare practitioners, patients, and care partners.

Have you ever considered what goes into building a lexicon? A lexicon is the vocabulary of a person, language, or branch of knowledge. It is an observed, recorded, and parsable dataset that can be leveraged in every category of marketing. The words and phrases that are used to describe a brand, including the technical and clinical story describing the mechanism of action, create the brand's lexicon. These words need to tell an accurate, differentiating story that the brand can own. The lexicon should also be both interesting enough to evoke the desired emotions and able to be sustained over time. The lexicon can also act as a guideline for a marketing program driving organic and paid search, help understand a consumer journey, and help define novice from expert in your audiences.

Although you may not yet be familiar with computational lexical analysis research, it may be just what your brand needs to be differentiated and emotive. This lexicon research can help as you build a scientific platform or narrative, or even help redefine a healthcare category. It can be leveraged to understand the differences in healthcare professional language as compared to patient and care partner language, or help develop campaign and training material for healthcare professionals to use with patients and their care partners.

For many brands, the language of a brand is a hallmark. Many of the modern channels of communication need to be salient and move back and forth between platforms and channels. Lexicon, when properly defined and integrated, can travel with the brand. It can be as effective as a logo or brand color to orient a consumer around what a brand can stand for.

We have a team of linguists, both computational linguists and sociolinguists by training, who conduct computational lexical analysis research. They first build a large corpus of data from published literature in the relevant category. These documents are then analyzed using a variety of tools and approaches, working closely with our medical experts and account teams to provide decision-making parameters around the lexicon and address the research questions. Then, we bring in clients to discuss the initial findings. This can be a small team, or larger, and include as many stakeholders as possible, such as brand managers, research managers, scientists, medical, even legal/regulatory. Based on this discussion, we may go back to the corpus and further analyze the content.

The corpus of data is often created by pulling all available articles that have been published in the PubMed Central Open Access Collection, which allows us access to use the content free of charge, (http://www.ncbi.nlm.nih.gob/pmc/tools/openftlist/) and fit specific search criteria. The broader, the better when it comes to building the corpus – and more refined searches can be done for analysis. For example, include 5-10 years of data, and search for general category publications. You can also add other published content to this database, to make it as robust as possible. We typically don't exclude content based on article type, journal, or country of origin (as long as content is in English for consistency). This will allow for hundreds to thousands of articles, depending on the criteria, thousands of authors, and millions of words to be analyzed – and a broad enough look at the content to examine the language used, without bias.

It is this process that allows for a truly scientific perspective on defining lexicon. Because the source data is large enough in scope, the computational analysis can eliminate personal, demographic, and regional discrepancies resulting in an output to inform the next steps.

Once the corpus of data is built, we combine computational and qualitative approaches to conduct linguistic analyses. The facts are gathered and all key questions are reviewed in

Compliments of PharmaVOICE

(c) PharmaLinx LLC. Rights do not include promotional use.
For distribution or printing rights, contact mwalsh@pharmavoice.com

depth. We examine what words are related to each other and what they mean in the context in which they are being used. We can create smaller micro-datasets on a focused topic to help assess the emergent language within that subset of content. For example, you can explore inflammatory bowel disease as a larger category, and ulcerative colitis or Crohn's disease as a subset for meta-analyses. The end result is powerful, foundational research for your brand.

Research: Five Tips for Success

It may all sound great in theory, but if you are wondering what you can really do with this research? Here are the top 5 ways our clients have been using this research successfully:

1. **Build the scientific platform** for a new brand from which all messaging work will be created.
2. **Create a positioning statement** for either a new or existing brand that is undergoing a change, such as a new indication, a new formulation, or competition joining the market. It could look something like this:

> **Brand X offers description word that differentiates brand, which is most important feature that does this element which is special, and is the only new 'description' word here that description lexicon to go here among target.**

3. **Eliminate potential nomenclature collision** with other brands within a category. Redefine the ambiguous: How should the area that is below your face, under your chin region, but not considered part of the neck be defined? We looked at both chin and neck in a corpus of data.
   Both terms lacked specificity — demonstrated by frequent use of modifying terms Also, at times, each term, chin and neck, is distinctly differentiated from the submental region/area, which is a more specific term for the treatment area.
   Differentiating words: When looking at words such as attack and flare, they are often used in place of one another. In one dataset examined, patients did not show a preference for either attack or flare — but parallels from other medical conditions suggest that flare is preferred and does a better job of conveying the repetitive nature and implies that flares can return, a critical part of disease education for the category studied.
4. **Create a glossary** that clearly explains the meaning of words in a specific category and outlines what words to use and not use. These are simple examples of language used for demonstrative purposes. Actual glossary definitions may be just as simple, or much more complex, depending on the category.

Preferred Terminology: Language to use

Term	Rationale
Everyday or daily medication	▶ Language already recognized by patients ▶ Clearly instructs them on how often they will need to take their medication
Preventative medication	▶ Not currently in patient lexicon ▶ Patient-friendly language ▶ Easily combined with everyday or daily to provide clear directive language (e.g., take this medication daily)

Non-preferred Terminology: Language to avoid

Term	Rationale
Clearance (in a respiratory category)	▶ Not patient-friendly terminology

5. **Redefine a market category** that is not clearly defined so it can be linked directly to your brand.
   For healthcare categories that are not clearly defined, either because they have been around a long time or possibly because there are new MOAs entering the market — redefining a category provides an opportunity to change things up. Think of how language has evolved in gout. The same word is used to describe both the disease and the acute flares that people experience. Depression and erectile dysfunction were previously conditions that were not discussed, but now, with refined language, think about how much they have changed.

The examples noted above are only a few that illustrate what can be done when focusing on lexicon for modern communications. They are also examples of how to leverage content within a category to understand opportunities, competitive challenges, audience preferences, and market activities to inform decision making. Equally important, as with any data you may use to inform your opinion, is understanding how to benchmark insights to recognize market changes and evolutions. In the case of a computational lexical analysis, a brand may choose to look for changes that occur over time and the length of time required for a specific disease category to adopt a new or modified vernacular. This is especially interesting when applied against market data. What are the most impactful channels to affect change in scientific exchanges? Who are the audiences that can effect the most change?

The practice, as it is presented, is completely platform-agnostic, but digital media is especially well-suited to monitor these changes through the collection of data, continued analysis of lexicon in publications, social media listening, search engine performance, and online surveys.

Because text offers such a large yield with regard to data, it becomes critical to evaluate program responses and the insights provided by our agency. There is an opportunity for every brand to begin the development of a lexicon and manage its evolution for use as a point of differentiation and foundation for the modern healthcare brand. Other ethnographic research methodologies such as in-home or in-office observational research can be conducted to add depth and validate materials developed through computational lexical analysis findings. These kinds of research methodologies are less traditional, but directly assess the emotional impact.

Still wondering if this is right for your brand?

A computational lexical analysis produces scientifically based findings that can enhance the language and improve overall messaging and discourse across all avenues of communication. It can be done in a relatively quick timeframe, and can provide insights that are important to all stakeholders. PV

Ogilvy CommonHealth Worldwide, *part of WPP Health & Wellness, committed to creativity and effectiveness in healthcare communications, everywhere.* For more information, visit ogilvychww.com.

(c) PharmaLinx LLC. Rights do not include promotional use.
For distribution or printing rights, contact mwalsh@pharmavoice.com

Compliments of PharmaVOICE