

# How We Talk About People with Conditions: Trends Over Time Using Google Ngrams

Andrew W Brown

Office of Energetics and Nutrition Obesity Research Center, University of Alabama at Birmingham

## Background

- People first language (PFL) is the practice of referring to a person separately from a health- or physical-condition he or she may have.
- Some advocate for using PFL to increase the respect given to individuals<sup>1</sup>
- *Diabetes*, in the guide to authors, states: “The term *diabetic* should not be used as a noun.”<sup>2</sup>
- However, PFL can make for ‘horrible prose,’<sup>3</sup> and the use of clinically relevant terms can sometimes be seen as stigmatizing.<sup>4</sup>
- Some communities take ownership of their condition, such as the capitalization of Deaf for the Deaf community.<sup>5</sup>
- The purpose of this study was to investigate trends in the usage of different types of language referring to individuals with four conditions over time in published material included in the Google Books Ngram Viewer (GBNV) project.

## Methods

Four conditions with common noun and adjective forms were chosen:

Noun	Adjective
Asthma	Asthmatic
Autism	Autistic
Diabetes	Diabetic
Obesity	Obese

Identified 5 types of language describing people with conditions:

Language	Structure	Example Sentence
1) Defining	Condition-Noun	<i>Asthmatics</i> rely on inhalers.
2) Adjective	Condition-adjective_subject	The <i>autistic child</i> did not make eye contact.
3) PFL Adjective	Subject <i>is</i> condition-adjective	The man who <i>is obese</i> is working out.
4) PFL Neutral	Subject <i>with</i> condition-noun	The woman <i>with diabetes</i> injects insulin.
5) PFL Negative	Subject <i>negative verb</i> with condition-noun	The employee <i>suffers from asthma</i> .

Search GBNV for each type:

**Type 1)** Search GBNV for *condition\_NOUN* (1-grams).

**Type 2)**

- Get all 2-grams from the “as”, “au”, “di”, and “ob” datasets from GBNV.
- Identify all 2-grams starting with the adjective form of the conditions.
- Identify all person-related nouns for the second word in the 2-grams.
- Extract all 2-grams with adjective-forms and person-nouns.

**Type 3)** Use all 2-grams from datasets in Type 2 that do not have adjectives or nouns as parts-of-speech following the adjective-form of the condition (e.g., “diabetic\_ADJ went”).

**Type 4)** Searched for 3-grams: noun; “with” or inflections of “has”; the condition (e.g. “\_NOUN\_ has obesity”). Phrases starting with animal nouns were subtracted (e.g., “dog with diabetes”).

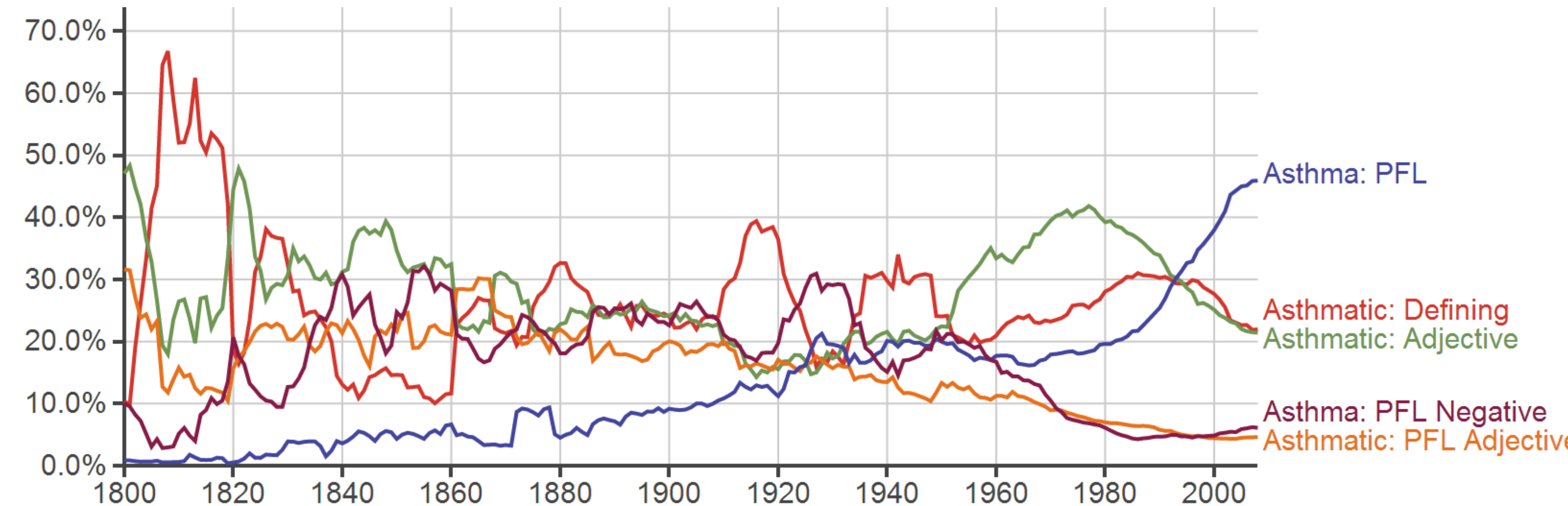
**Type 5)** Searched for 3-grams of phrases such as “victim of”, “suffering from”, and “tormented by” preceding the condition-noun.

## Results

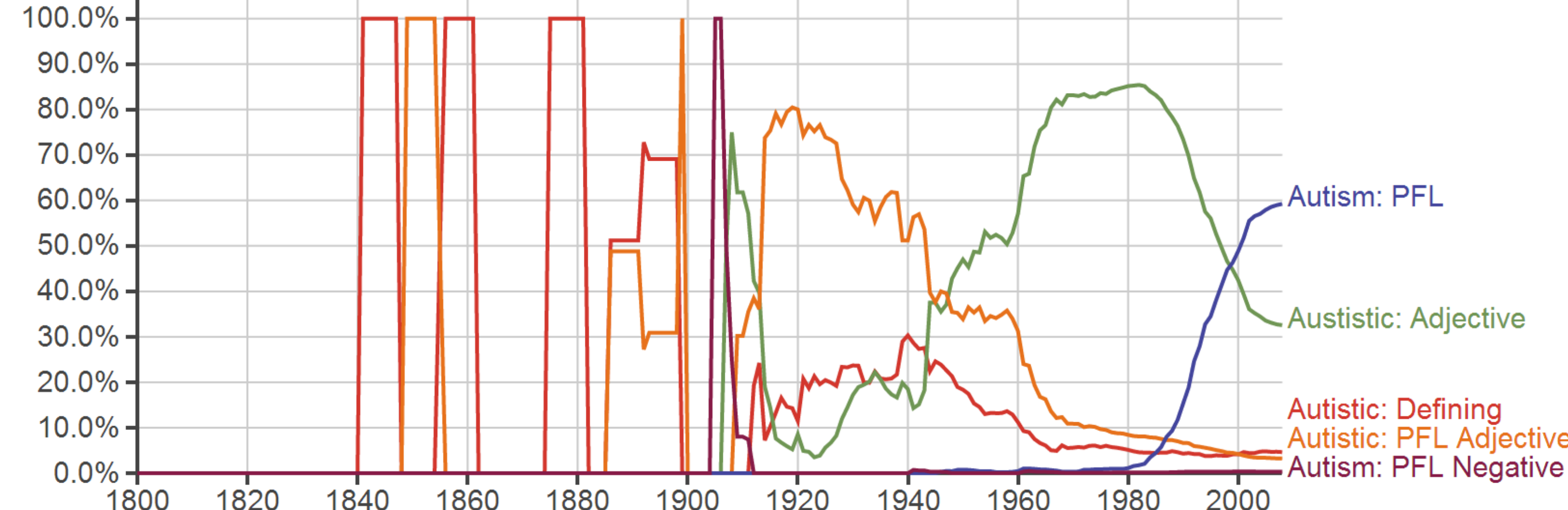
	Noun	Adjective	Other
Asthma	678799	28506	61206
Asthmatic	241	63630	2
Asthmatics	22518	28	40
Autism	602910	3436	4581
Autistic	8042	205848	638
Autistics	4411	0	51
Diabetic	12098	459957	448
Diabetics	76229	64	12
Diabetes	1730875	23616	84350
Obese	1962	331141	7772
Obesity	805370	35552	29478

**Table: Parts of speech for iterations of words for four investigated conditions.** GBNV was searched for the relative proportions of each part of speech in 2008. Filled black circles indicate the highest frequency within a word (e.g., Asthma is most frequently referred to in noun form). Shading represents relative contribution to total references across all words (e.g., Diabetes as a noun was most frequently cataloged). Numbers represent billionths of a percent of all cataloged words.

**Figure 1a: Percentage of each type of language used to describe people with ASTHMA**



**Figure 1b: Percentage of each type of language used to describe people with AUTISM**



**Figure 1c: Percentage of each type of language used to describe people with DIABETES**



**Figure 1d: Percentage of each type of language used to describe people with OBESITY**



The y-axis represents the relative percent of each type of language used within each condition. The x-axis represents calendar years. In Figure 1b, the stark 0-100% lines are an artifact of few references to the condition being made in those particular years (see Figure 3). Lines are smoothed over 3 years.

Interactive Graphs



tinyurl.com/PFLcond

## Results

**Figure 2: Trends in descriptions of people and their conditions over time**



The y-axis estimates the percentage of all ngrams dedicated to describing people with these four conditions, pooled across the 5 identified types of descriptive language. Data points are smoothed over 3 years.

Interactive Graph



tinyurl.com/PFLabs

## Discussion and Conclusions

Language use varies greatly from condition to condition and across time.

Research should be conducted on the receptivity of patients to how they are addressed, but most people can agree that respect for individuals should be at the center of health-condition communication.

Limitations:

- GBNV only includes materials in Google Books
- Ngrams depend on proper parts-of-speech tagging<sup>6,7</sup>
- This analysis is dependent on correct identification of relevant ngrams (e.g., exhaustive set of people-describing nouns)
- It mixes ngrams of different sizes (1-grams, 2-grams, and 3-grams)
- Rare phrases may be missed (only ngrams that occur 30 times or more are included in the corpus)
- Does not include euphemisms, synonyms, or pejorative versions of conditions

## References

1. Guidelines: How to write and report about people with disabilities, 8<sup>th</sup> edition. [rtcil.org/guidelines.shtml](http://rtcil.org/guidelines.shtml)
2. *Diabetes* Information for Authors: <http://diabetes.diabetesjournals.org/site/misc/ifora.xhtml>
3. Collier R. *Canadian Medical Association Journal*. 2012
4. Puhl R. *International Journal of Obesity*. 2012
5. National Association of the Deaf: <http://www.nad.org/issues/american-sign-language/community-and-culture-faq>
6. Michel et al. *Science*. 2010
7. Lin et al. *Proceedings of the 50<sup>th</sup> Annual Meeting of the Association for Computational Linguistics*

## Acknowledgments

AWB is supported by T32 DK62710

