
Challenges for multilingual metaphor corpora: Methods and guiding principles

Elise Stickles¹, Schuyler Laparle², Celeste Browning¹,
Vivian Du², Kelly Jones^{2,3}, Amine Lahouli^{4,5},
Inés Lozano⁶, and Eve Sweetser²

¹ University of British Columbia – Vancouver, Canada

² University of California – Berkeley, USA

³ University of California – San Diego, USA

⁴ Université Sorbonne-Nouvelle Paris 3, France

⁵ Carthage University, Tunisia

⁶ Universitat Politècnica de València, Spain

RaAM September 21, 2022

Overview

- Background:
 - Why build a multilingual metaphor database?
 - History and current status of the project
- Methods for building a multilingual database
 - Bottom up vs. top down data collection and analysis
 - Challenges we've encountered
- Guiding principles for building a multilingual database
 - Aim to navigate and possibly avoid challenges

Motivations for a multilingual metaphor database

- Increasing interest in computational approaches to automatically identifying metaphor in text (e.g. Veale et al. 2016; Klebanov et al. 2018; Gangemi et al. 2018)
- Simultaneously, corpus approaches to metaphor research have been enabled by technological advancements, and encouraged by the quantitative turn in cognitive linguistics
- The result is a need for ontologically informed, structured databases useful for computational and corpus research in figurative language (Bolognesi et al., 2019; Olza et al. 2021)

History of the current project

- MetaNet began in 2012 as a government-funded project in automated metaphor identification and analysis (Dodge et al. 2015)
- Creation of MetaNet database of structured metaphors and frames (Stickles et al. 2016a)
 - Foundational work on primary metaphors (Grady 1997)
 - Culturally specific metaphors (Kövecses 2005, David et al. 2016)
 - Focus on metaphors related to social issues such as poverty (Dodge 2016), drug abuse (Stickles et al. 2016b), gun control (David et al. 2016)
 - Transitioned to analysing metaphors for cancer (e.g. Sweetser & Laparle 2019) and now COVID-19 (e.g. Sweetser et al. 2021)

Status of the current project

- Currently housed at the University of British Columbia - public website coming April 2023
- Expanded to cover multiple languages and varieties:
 - American and Canadian English
 - American and Mexican Spanish
 - Canadian and Hexagonal French
 - Mandarin Chinese
- Actively adding example data, new frames and metaphors to the database
 - Current focus on metaphors for cancer, COVID-19, and climate change
 - See our other RaAM talks for examples of our current COVID-19 metaphor studies:
 - Laparle et al. on moral framing in pandemic metaphors, today at 17:35
 - Browning & Stickles on Canadian vs. American metaphors, today at 18:05

Methods and Challenges: 3 approaches to corpus work

Stefanowitsch (2008) suggest three basic ways to approach corpus-based metaphor analysis:

1. Search for the **Source domain** (we call this **top-down**; cf. Deignan 2005)
2. Search for the **Target domain** (we call this **bottom-up**; cf. Deignan 2005)
3. Search for both! (we extend this to an **iterative** approach)

All three are important in building a comprehensive multilingual corpus

Methods and Challenges: top-down or bottom-up?

- WAR and JOURNEY metaphors are prolific in discourses on DISEASE
- Knowing this, a huge body of work has been done looking at the use and effectiveness of these metaphors
 - CANCER: e.g. Harrington 2012; Hendricks et al. 2019; Landau et al. 2018; Marron et al. 2020; Semino et al. 2017...
 - COVID: e.g. Marron et al. 2020; Panzeri et al. 2021; Sabucedo et al. 2020; Wicke & Bolognesi 2020...
- This is a **top-down** approach to metaphor research – we have a particular source domain in mind, and we go out to see how it is used.

Methods and Challenges: top-down or bottom-up?

- If only a top-down approach is used when corpus building, you are going to miss *a lot* of cool metaphors by missing other source domains.
- In our corpus building we practice **top-down** and **bottom-up** methodologies iteratively
 - bottom-up methodologies help us identify new source domains
 - top-down methodologies with a team of native speakers help to facilitate the finding and documenting of less frequent mappings from a source domain across language variants
- Let's walk through what this looks like...

Methods and Challenges: bottom-up data collection

- Individual members are responsible for gathering data from a particular genre (e.g. news, blogs, scientific articles) in a particular language variety
- Each text is then analyzed for all metaphors present and brought to the group for discussion (similar to MIPVU: Steen 2010)

Methods and Challenges: bottom-up data analysis

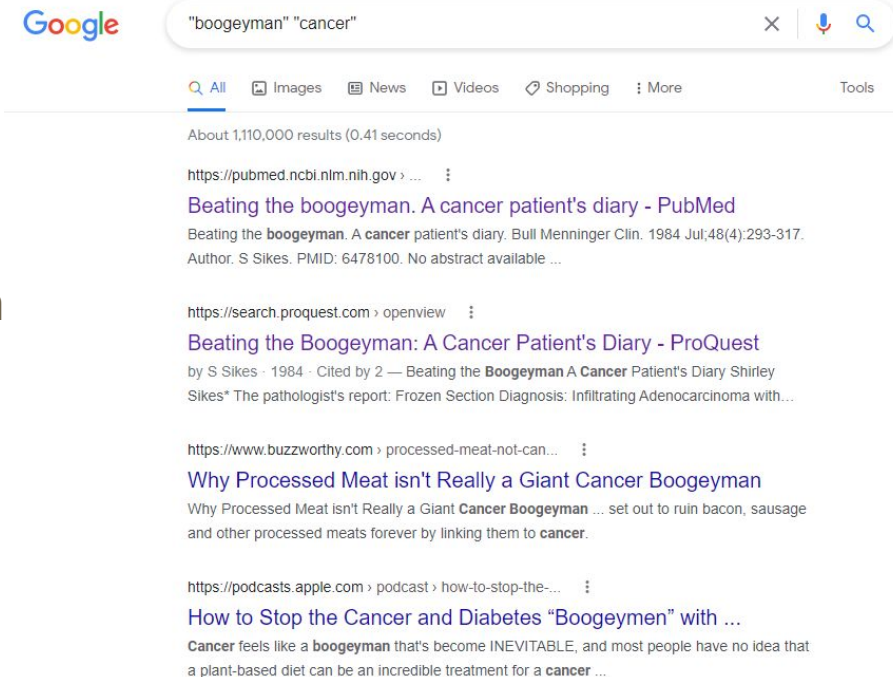
- The annotated text is brought to the group. Novel metaphors are discussed and terminology for domains and mappings are agreed upon
- The new domains, mappings, and lexemes then added to our database



year ago, Joe Biden **launched his “cancer moonshot,”** a major **national push** to improve the prevention, detection, and treatment of cancer, a plan that was widely recognized to be incremental. “I believe that we need an absolute national commitment to end cancer as we know it,” Biden said while he was on his tour to cancer centers at Penn and Duke University. “I’m not naïve. I didn’t think we could ‘end cancer.’ I’m not looking for a **silver bullet.** There is none.” Many thought the “moonshot” risked casting the solution to cancer as an engineering problem.

Methods and Challenges: top-down data collection

- We then use our newly analyzed metaphors to brainstorm possible related metaphors and likely lexemes across language variants. This is where top-down is helpful again.
 - E.g. We know from our analysis of one monster metaphor that **BOOGEYMAN** likely has similar mappings



A screenshot of a Google search for "boogeyman" cancer. The search bar shows the query "boogeyman" cancer. Below the search bar, there are navigation options: All, Images, News, Videos, Shopping, and More. The search results show approximately 1,110,000 results in 0.41 seconds. The first result is from PubMed: "Beating the boogeyman. A cancer patient's diary - PubMed". The second result is from ProQuest: "Beating the Boogeyman: A Cancer Patient's Diary - ProQuest". The third result is from Buzzworthy: "Why Processed Meat isn't Really a Giant Cancer Boogeyman". The fourth result is from Apple Podcasts: "How to Stop the Cancer and Diabetes 'Boogeymen' with ...".

Methods and Challenges: top-down data collection

- We then use our newly analyzed metaphors to brainstorm possible related metaphors and likely related lexemes. This is where top-down is helpful again.
 - E.g. We know from our analysis of one monster metaphor that **BOOGEYMAN** likely has similar mappings
 - But that is likely *not* the case for **SASQUATCH**

The screenshot shows a Google search for "sasquatch cancer". The search bar contains the text "sasquatch" "cancer". Below the search bar, there are navigation tabs for All, News, Images, Shopping, Videos, and More. The search results show approximately 1,630,000 results in 0.37 seconds. The "Images" tab is selected, displaying a grid of images related to "sasquatch cancer". The images include a pink Sasquatch illustration, a black t-shirt with a Sasquatch graphic, a black hoodie with a Sasquatch graphic, and a black notebook with a Sasquatch graphic. Below the image grid, there is a "View all" button and a "Feedback" link. The search results also include product listings for "Breast Cancer Journal: Bigfoot Sasquatch Funny Pink ..." on Amazon.com. The listings show the journal's cover, price (\$4.99), and availability (in stock). The journal is described as a "Pink Ruled Lined Notebook - Diary, Writing, Notes, Gratitude, Goal Journal ...". The publisher is "Independently published (Au...".

Guiding Principles

- Guiding principles in response to challenges:
 1. After initial exploratory research, conduct in-depth language analyses independently;
 2. Resist treating one language as dominant in corpus construction and organization;
 3. Consult native speakers throughout.

Guiding Principles

1. After initial exploratory research, conduct in-depth language analyses independently
 - Read data for ALL metaphor, not just what one is expecting
 - Individual compilation and identification of metaphors
 - Study on identified metaphors and related metaphors in context
 - Collective discussion of examples

Summary of methods

- Bottom-up approaches are necessary for discovering the *breadth* of metaphoric expression (i.e. Source domain variation)
- Top-down approaches are necessary for understanding the *depth* of metaphoric expression (i.e. lexical and mapping variation)
- Both are necessary, especially where multiple language variants are considered. Each analyst brings a unique *linguistic, cultural* and *experiential* expertise to the project.
- The strength of an iterative approach is in using the diversity of expertise to turn a bottom-up discovery into careful and informed top-down excavations

Guiding Principles

2. Resist treating one language as dominant in corpus construction and organization
 - If your lingua franca is English... how do you prevent biases toward English-based metaphor searches?
 - Multilingual search (e.g. varieties of French, Spanish, Chinese, English)
 - Collective work is shared in sessions to provide points of contrast between different languages

Guiding Principles

3. Consult native speakers throughout

- Recognizing imperfect lemma translation
 - Animal metaphors are often very similar across languages, except in the exact *identity* of the animal
 - *scaredy cat* meaning ‘coward’ in English corresponds to *poule mouillée* “wet hen” in French
 - BODY IS A CONTAINER FOR THE EMOTIONS is quite common cross-linguistically, but which organ “contains” a particular emotion varies:
 - compare Mandarin *Ta pi-qi hen da* lit. ‘He’s got big gas in spleen’, meaning “He’s hot-tempered” (Yu 1995) to Spanish *Me recome los hígados* lit. ‘it’s eating my livers’ meaning “it makes me angry”.
- Consulting native speakers also helps adjust the meaning nuances of metaphors
- Multilingual composition of the group allowing feedback

References

- Bolognesi, M., Brdar, M., & Despot, K. (Eds.). (2019). *Metaphor and metonymy in the digital age: Theory and methods for building repositories of figurative language*. John Benjamins Publishing Company.
- David, O., Lakoff, G., & Stickles, E. (2016). Cascades in metaphor and grammar: A case study of metaphors in the gun debate. *Constructions and Frames*, 8(2), 214-255.
- Deignan, A. (2005). *Metaphor and corpus linguistics* (Vol. 6). John Benjamins Publishing.
- Dodge, E., Hong, J., & Stickles, E. (2015). MetaNet: Deep semantic automatic metaphor analysis. In E. Shutova, B. B. Klebanov, & P. Lichtenstein (Eds.). *Proceedings of NAACL 2015 Workshop on Metaphor*, pp. 40-49.
- Gangemi, A., Alam, M., & Presutti, V. (2018). Amnestic forgery: An ontology of conceptual metaphors. arXiv preprint *arXiv:1805.12115*.
- Grady, J. E. (1997). *Foundations of meaning: Primary metaphors and primary scenes*. PhD Dissertation, University of California, Berkeley.
- Harrington, K. J. (2012). The use of metaphor in discourse about cancer: a review of the literature. *Clinical journal of oncology nursing*, 16(4), 408-412.
- Hendricks, R. K., Demjén, Z., Semino, E., & Boroditsky, L. (2018). Emotional implications of metaphor: Consequences of metaphor framing for mindset about cancer. *Metaphor and Symbol*, 33(4), 267-279.
- Klebanov, B. B., Shutova, E., Lichtenstein, P., Muresan, S., & Wee, C. (Eds.) (2018). *Proceedings of the workshop on figurative language processing*. Association for Computational Linguistics, New Orleans, Louisiana.
- Kövecses, Z. (2020). *Extended conceptual metaphor theory*. Cambridge University Press.
- Landau, M. J., Arndt, J., & Cameron, L. D. (2018). Do metaphors in health messages work? Exploring emotional and cognitive factors. *Journal of Experimental Social Psychology*, 74, 135-149.
- Marron, J. M., Dizon, D. S., Symington, B., Thompson, M. A., & Rosenberg, A. R. (2020). Waging war on war metaphors in cancer and COVID-19. *JCO Oncology Practice*, 16(10), 624-627.

References

- Olza, I., Koller, V., Ibarretxe-Antuñano, I., Pérez-Sobrino, P., & Semino, E. (2021). The# ReframeCovid initiative: From Twitter to society via metaphor. *Metaphor and the Social World*, 11(1), 98-120.
- Panzeri, F., Di Paola, S., & Domaneschi, F. (2021). Does the COVID-19 war metaphor influence reasoning?. *PloS one*, 16(4), e0250651.
- Semino, E., Demjén, Z., Demmen, J., Koller, V., Payne, S., Hardie, A., & Rayson, P. (2017). The online use of Violence and Journey metaphors by patients with cancer, as compared with health professionals: a mixed methods study. *BMJ supportive & palliative care*, 7(1), 60-66.
- Steen, G. (Ed.). (2010). A method for linguistic metaphor identification: From MIP to MIPVU. John Benjamins Publishing.
- Stefanowitsch, A. (2008). Corpus-based approaches to metaphor and metonymy. In A. Stefanowitsch & S. Th. Gries (Eds.), *Corpus-based approaches to metaphor and metonymy* (pp. 1-16). Berlin, NY: De Gruyter Mouton. doi: <https://doi.org/10.1515/9783110199895.1>.
- Stickles, E., David, O., Dodge, E. K., & Hong, J. (2016a). Formalizing contemporary conceptual metaphor theory: A structured repository for metaphor analysis. *Constructions and Frames*, 8(2), 166-213.
- Stickles, E., David, O., & Sweetser, E. (2016b). Grammatical constructions, frame structure, and metonymy: Their contributions to metaphor computation. *Proceedings of the 11th Meeting of the High Desert Linguistics Society (HDLS) 2015*, pp. 317-345.
- Sweetser, E. & Laparle, S. (2019) War is war - or is it? Different genres show different metaphors for cancer. Talk presented at *15th International Cognitive Linguistics Conference*. Kwansai Gakuin University, Nishinomiya, Japan.
- Sweetser, E., Laparle, S., Jones, K., Getter, D., & Kulgold, A. (2021). Choosing source domains to guide metaphoric reasoning. Talk presented at *Researching and Applying Metaphor*, Vilnius, Lithuania.
- Wicke, P., & Bolognesi, M. M. (2020). Framing COVID-19: How we conceptualize and discuss the pandemic on Twitter. *PloS one*, 15(9), e0240010.
- Veale, T., Shutova, E., & Klebanov, B. B. (2016). *Metaphor: A Computational Perspective*. Morgan & Claypool.
- Yu, N. (1995). Metaphorical expressions of anger and happiness in English and Chinese. *Metaphor and symbol*, 10(2), 59-92.

Thank you!