**ALGEE-HEWITT — Unspoken**

<2 figures>

**The Unspoken Word: Race and the New Language of Identity**

Algee-Hewitt, B. and Algee-Hewitt, M.

In a statement on the meaning of ‘race’ within physical anthropology, Alice Brues (2009) raises the spectre of a mismatch between our biological understanding of human variation, and the work that the word ‘race’ performs within academic, and specifically anthropological, literature: ‘The visible difference between different populations of the world tells everyone that there is something there. We had better be prepared to explain what is there, and why, before we discuss what it does or does not mean’*.* Yet contrary to the meaning behind her question, it is apparent that race *already* has meaning as a clearly articulated concept in both the popular and academic understanding, whether it exists in an objective biological reality or not. Yet despite its ubiquity, few discourses are more contentious. At best, critics claim that race is cultural construction with no basis in biological fact; at worst, the language of ancestry has been accused of perpetuating the dark histories of racial categorization and eugenics. But our understanding of the evolution of this discourse is clouded both by individual agendas and specific cultural concerns: what we lack, we argue, is a comprehensive and critical understanding of the patterns that the language of race has assumed throughout recent academic discourse, and specifically within the discipline of physical anthropology, whose practitioners are, whether consciously or unconsciously, implicated in this discourse through their work on human variation and forensic identification.

Through this paper, we seek to explore the latent discourse of race that characterizes the anthropological study of human variation. How has the emergence of forensic anthropology and genetics altered the way in which physical anthropologists interpret concepts such as race, ancestry, or ethnicity? Are there discrete and recognizable patterns of discourse that have shaped our understanding of race? This project aims to reconstruct this history by bringing to bear the new methodologies of digital humanities on a representative sample of scholarly articles on ‘race’ from the field of anthropology. The goal is not to reify race as a concept, but explore how the discourse is operational within the work of anthropology, considering both its literature that supports it and literature that argues against it. By leveraging the new computational techniques of the digital humanities, we seek to isolate the ways in which the language surrounding the concept of race has changed over the last century and how these discursive shifts have organized the approach to human variation in physical anthropology. We provide comparative data by also exploring the same questions for cultural anthropology and the forensic sciences.

The way we talk about race, in other words, is an indicator of how we understand it.

**Sample**

As our study is primarily concerned with the ongoing transformation of ‘race’ in the field of physical anthropology, the majority of our sample is taken from articles published in the *American Journal of Physical Anthropology* (AJPA) throughout its publication history. Similarly, as current anthropological thinking on race, whether as a proxy for biological difference or as a social reality, is deeply affected by the practical necessities of the forensic sciences, we also drew part of our sample from articles published in the *Journal of Forensic Sciences* (*JFS*). In order to compare the specific configurations of ‘race’ within the discourse of physical anthropology to the study of the field at large, the final subset of our sample is composed of articles published in *Current Anthropology* (*CA*). In addition to these three core publications, we also add to our study a corpus of genetics articles that allow us to capture how the language of race is inflected by recent work in human genetics. For this part of our corpus we draw on the journals *Human Biology*, the *American Journal of Human Genetics*, and *Genetics*, using the same search criteria as in our physical anthropology corpus. By combining peer-reviewed academic articles from these six publications, our study is able to examine the transformations in the anthropological discourse of race through time and across the disciplinary configurations of the field.

Our primary physical anthropological sample contains 440 articles published between 1918 and 2013 and that contain some cognate of the words ‘race’, ‘ancestry’, or ‘ethnic’ within either the title or article description. Of these, 299 were published in *AJPA*, 79 were published in *CA*, and 63 were published in *JFS*. In preparation for our analysis, we divided our sample twice: by publication and by period. These two taxonomies of the corpus allowed us to identify trends in discourse over time (period) and across the disciplinary boundaries (publication). The period divisions were informed by the distribution of the sample and correspond to the early to mid-20th century (1918–1972 or Early), the late 20th century (1973–1999 or Mid), and the 21st century (2000–2013 or Late). Our turn towards genetics has allowed us to more than double our corpus size and gives a critical window into the ways in which the discourse around recent technological innovation is bringing back older conceptions of human variation

**Methodology**

*Distinctive Words*

While unsupervised computational linguistic techniques such as topic modeling (Blei, 2006; 2007) would allow us to extract clusters of highly correlated words from across the sample (thus approximating the ‘topics’ of the articles), the variety of these clusters would obscure the nuances of any potential ‘race’ topic. In other words, we seek to capture the ways that the discourse of race manifests itself even (and especially) when the topic of race is not overtly at issue. Instead, therefore, we chose to target specifically the language of ‘race’ and to extract the vocabulary in which it is embedded. We queried our sample of texts for our target terms (‘race’, ‘ancestry’, ‘ethnic’) and extracted the 15 words before and after, effectively performing a collocation analysis on each text. We selected 15 as our horizon as it is the average length of a sentence within our sample, ensuring that each extraction captured at least one complete syntactical unit. After excluding both stopwords and low frequency words (n<50), we applied a Fisher’s Exact Test to each of the remaining words to test for significant word usage (α=0.05) within the vicinity of either ‘race/racial’, ‘ancestry/ies’, or ‘ethnic/ity,’ with the goal of identifying the category of racial terminology with which each word was most highly correlated. These sets of significant words (n=230) represent the lexical ‘fingerprint’ of each of our concepts.

By tracing the relationships between articles based on their usage of these words, we are able to identify how the discussion of race structures the corpus across time and disciplinary approach. Similarly, we are also able to trace the relationships between the words themselves in order to uncover the ways in which these individual discourses co-evolved over the course of the 20th century.

*Topologies*

Our primary visualization strategy uses a topological representation of either the corpus (Figure 1) or the semantic fields (Figure 2). Based upon an extension of a visualization technique using Voronoi diagrams developed by Bell Labs (Gansner et al., 2009; Algee-Hewitt and Piper, 2013; Algee-Hewitt, 2011; Okabe et al., 2000), the topology reconfigures similarity as a function of distance within a spatial representation. The resulting map provides a more complex visualization of relationships among the individual members and their local neighborhoods, replacing raw distances and word frequencies with a depiction of relationality based on proximity, area, and conjunction. The close proximity of article-tiles in our corpus maps indicates a strong similarity in their use of the vocabulary of race, while articles located on opposite sides of the topology indicate their difference. Similarly, words appearing next to each other in the word maps appear frequently together, while those far apart are rarely used within the same text. Tile size within these maps is a function of density. The smaller the tile, the more tightly embedded each object is within its surrounding cluster, while large tiles indicate their dissimilarity from the immediate neighborhood. We are also able to visualize another layer of data by adding a color mapping to the topology.

Our corpus maps are colored for both period and publication: this reveals a strong disciplinary cohesion between the publication and the precise mix of words that characterize its articulation of the ‘race’ concept. Similarly, there is a less strong, but still significant correlation between the language of race in each article and its data of publication. Our word maps are colored based upon the individual word memberships in our race concept categories. This allows us to observe clusters of words that form across time and within particular disciplines, suggesting the ways in which the discourse of race has been affected by (and, in turn, shaped) the individual disciplines or study (physical anthropology, cultural anthropology, forensic anthropology, and genetics). Further, it allows us to witness the configurations of this discourse across our three period divisions: from the early studies of human diversity in the early 20th century through the current genetics revolution. Through a careful parsing of these topologies, we are able, in this paper, to give a clear and concise account of the ways in which the race concept has been mediated within the different disciplinary discourses of anthropology and, most importantly, the ways in which its evolving central term (from ‘race’ to ‘ethnicity’ to ‘ancestry’) has carried along with it discursive configurations that remain constant across all of our disciplines and periods.

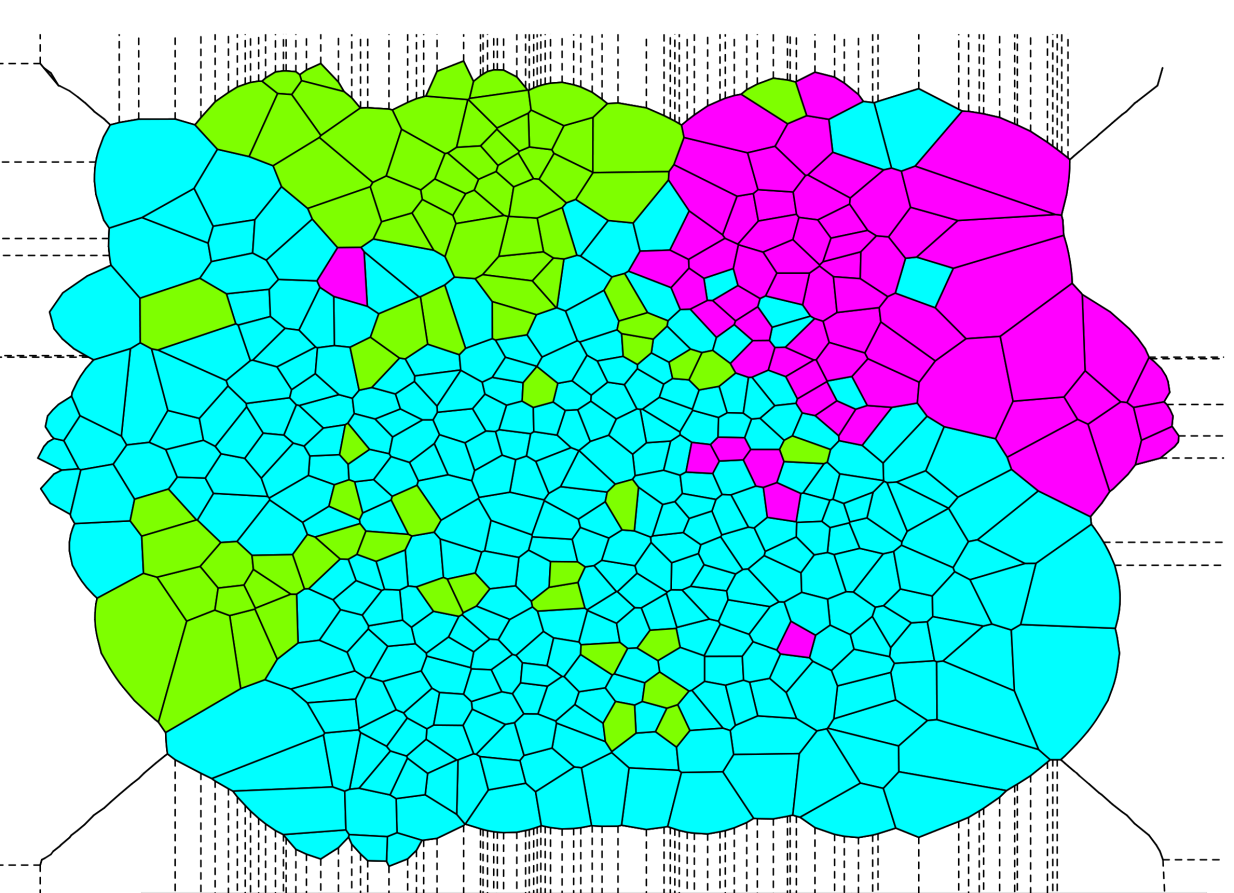
****

Figure 1. Topology of articles derived from shared clusters of words, colored for publication source.

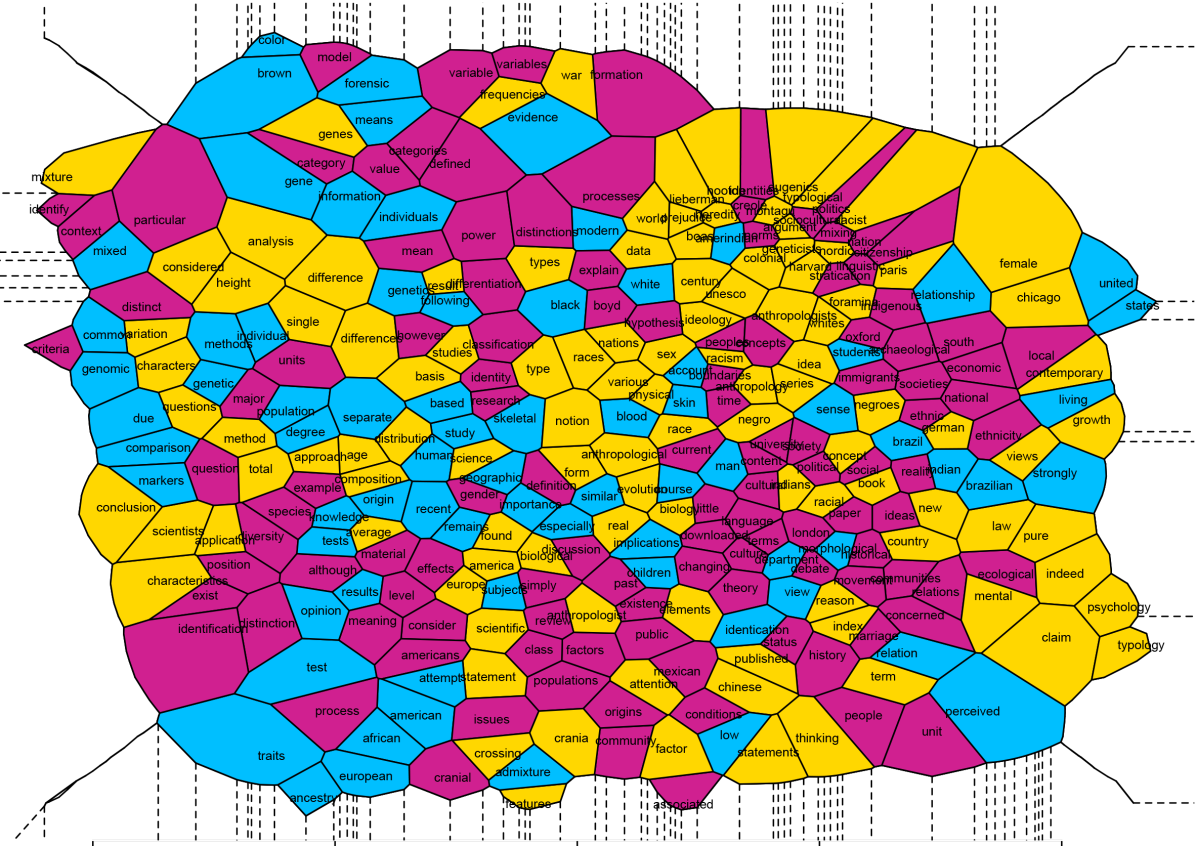


Figure 2. Topology of ‘race’ terms, clustered by co-occurrence, colored by term group membership.

**References**

**Algee-Hewitt, B. F. B.** (2011). *If and How Many ‘Races’? The Application of Mixture Modeling to World‐Wide Craniometric Variation.* Knoxville: TRACE, University of Tennessee.

**Algee-Hewitt, M. A. and Piper, A.** (2013). The Werther Effect I: Goethe Topologically. In Erlin M. and Tatlock, L. (eds), *Distant Readings/Descriptive Turns: Topologies of German Culture in the Long Nineteenth Century*. Rochester, NY: Camden House, 2014.

**American Anthropological Association (AAA)**. (1999). American Anthropological Association Statement on Race. *American Anthropologist,* **100**: 712–13.

**Blei, D. and Lafferty, J.** (2006). Correlated Topic Models. *Advances in Neural Information Processing Systems,* **18**: 147–54.

**Blei, D. and Lafferty, J.** (2007). A Correlated Topic Model of *Science*. *Annals of Applied Statistics,* **1**: 17–35.

**Brown, R. A. and Armelagos, G. J.** (2001). Apportionment of Racial Diversity: A Review. *Evolutionary Anthropology,* **10**: 34–40.

**Brues, A. M.** (2009). The Objective View of Race. In Gordon, C. C. (ed.), *Race, Ethnicity, and Applied Bioanthropology*. Chichester: John Wiley, pp. 74–78

**Cartmill, M.** (1998). The Status of the Race Concept in Physical Anthropology. *American Anthropologist,* **100**: 651–60.

**Foster, J. W. and Sharp, R .R.** (2002). Race, Ethnicity, and Genomics: Social Classifications as Proxies of Biological Heterogeneity. *Genome Research,* **12**: 844–50.

**Gansner, E., Hu, Y., Kobourov, S. and Volinsky, C.** (2009). Putting Recommendations on the Map—Visualizing Clusters and Relations. *Proceedings of the Third ACM Conference on Recommender Systems*, New York, pp. 345–48.

**Kaszycka, K. A., Štrkalj, G. and Strzałko, J.** (2009). Current Views of European Anthropologists on Race: Influence of Educational and Ideological Background. *American Anthropologist,* **111**: 43–56.

**Lieberman, L., Hampton, R. E., Littlefield, A. and Hallead, G.** (1992). Race in Biology and Anthropology: A Study of College Texts and Professors. *Journal of Research in Science Teachng,* **29**(3): 301–21.

**Okabe, A., Boots, B., Sugihara, K. and Chiu, S. N.** (2000). *Spatial Tesselations: Concepts and Applications of Voronoi Diagrams.* 2nd ed. Wiley, Chichester.

**Ousley, S. D., Jantz, R. L. and Freid, D.** (2009). Understanding Race and Human Variation: Why Forensic Anthropologists Are Good at Identifying Race. *American Journal of Physical Anthropology,* **139**: 68–76.

**Sauer, N. J. (**1992). Forensic Anthropology and the Concept of Race: If Races Don’t Exist, Why Are Forensic Anthropologists So Good at Identifying Them? *Social Science and Medicine,* **34**: 107–11.