Language Science Colloquium 2016-17

Monday, April 10th

Time: 11:00-12:30 pm
Location: Rm # 1321, Social and Behavioral Sciences Gateway Building
Lunch will be served.

Kimberly A. Jameson, Project Scientist,

Institute for Mathematical Behavioral Sciences, UC, Irvine

A color categorization resource for cross-cultural research: The ColCat Digital Archive

Abstract: The Mesoamerican Color Survey (MCS) collected color-naming and categorization data from approximately 900 speakers from 116 indigenous languages from regions in Mesoamerica. The MCS was developed, by Robert E. MacLaury, principle investigator of the survey, during the years 1978-1981, and analyses of these data were originally reported by MacLaury in the context of his Vantage Theory modeling approach.[1] The MCS is one of two large existing databases (the other being the World Color Survey, or WCS) that directly investigated, on a large scale, color naming and categorization across many linguistic societies. The MCS and WCS employ nearly identical standardized procedures for evaluating large numbers of color stimuli, languages and informants. It is estimated that more than 100 indigenous languages are spoken in Mexico and Central America. Like most languages each MCS language has a color lexicon that partitions environmental color appearance stimuli according to a pattern that is specifically relevant to a given language's speakers. Moreover, every MCS color categorization system also shares characteristics with systems observed for other Mesoamerican languages and with those of languages elsewhere in the world. Recognizing the value in the MCS data, the interdisciplinary ColCat research group at UC Irvine sought to convert the paper copy of MacLaury's research archive into a public access digital database, similar to that developed for the widely-known WCS.[2] The entirety of MacLaury's MCS data, as well as that from his multinational data collection efforts, is included in the ColCat digital archive, representing a total of 212 surveys. MacLaury's additional 96 color categorization surveys are, in their own right, valuable for their diversity in that they include native speakers from a wide variety of languages such as Slavic languages, Hungarian, Salishan languages of the Pacific Northwest United States, Zulu and several South Africa/ Zimbabwe languages, Native American languages, Germanic languages, European languages, and Asian languages. This talk presents research on preparing the archive's ~23,000 pages of paper copy into a public-access digital archive. In addition to manually transcribing a portion of the archive's handwritten data, we developed novel approaches for rapidly converting the archive to digital copy. Research results are presented for two complementary lines of investigation: the first involving approaches using Optical Character Recognition and machine learning to digitize the handwritten data, and a second approach using crowdsourced transcriptions of the data that were

aggregated via an innovative variation of Cultural Consensus Theory analysis.[3, 4] A further aim of the project was to develop a platform to make the archive available to the scientific research and teaching community. Towards this aim we present The Robert E. MacLaury Color Categorization (ColCat) Digital Archive website, containing all of MacLaury's data as a public-access color categorization digital archive. The ColCat platform[5], including features of the Graphical User Interface developed for organizing the archive as a web-based resource, is presented.

References:

[1] MacLaury, R. E., (1997). Color and Cognition in Mesoamerica: Constructing Categories as Vantages. University of Texas Press.

[2] Jameson, K. A., Benjamin, N. A., Chang, S., M., Deshpande, P. S., Gago, S., Harris, I. G., Jiao, Y., Tauber, S. (2015) "Mesoamerican Color Survey Digital Archive." Ronnier Luo (Ed.). Encyclopedia of Color Science and Technology, Volume 2. 909-921. Springer Science + Business Media New York. DOI: 10.1007/978-3-642-27851-8.

[3] Jameson, K. A., Deshpande, P. S., Tauber, S., Chang, S. M., Gago, S. (2016). Using individual differences to better determine normative responses from crowdsourced transcription tasks: An application to the R. E. MacLaury Color Categorization Archive. © 2016 Society for Imaging Science and Technology, 115.1-9. DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-115.

[4] Deshpande, P. S., Tauber, S., Chang, S. M., Gago, S., Jameson, K. A. (2016). Digitizing a Large Corpus of Handwritten Documents Using Crowdsourcing and Cultural Consensus Theory. International Journal of Internet Science.11.1 (2016): 8-32.

[5] Jameson, K. A., Gago, S., Deshpande, P.S., Benjamin, N.A., Chang, S.M., Tauber, S., Jiao, Y., Harris, I.G., Xiang, Z. Huynh, B.B., Ke, H., Lee, W.J., MacLaury, R.E. (2016). The Robert E. MacLaury Color Categorization (ColCat) Digital Archive. http://colcat.calit2.uci.edu/. The California Institute for Telecommunications and Information Technology (Calit2). University of California, Irvine. USA.

RSVP Patty Jones at: patty.jones@uci.edu