Results of previous NSF support: The current proposal is for renewed support of NSF award 1145762 (RUI: Conceptual Foundations of Language: $384,460, 6/15/12-5/31/16).

Intellectual Merit Contributions: The prior NSF project explored the nature of infants' event representations, asking: Do infants conceptualize 'goal' and 'source' broadly and abstractly, as they are represented in language? How does language acquisition modulate conceptualization? Since June 2012, 395 children (with 33% reporting to be non- Caucasian) were tested. The findings have been presented at 7 peer-reviewed conferences, published in *Language Learning and Development* and *Cognitive Science*, and as a book chapter in *Conceptual Change and Origins of Knowledge*. Further, two manuscripts have been invited for revision/resubmission to *Cognition* and *Infancy*. Intellectual merit is as follows: First, using a familiarization looking time study, as well as a Preferential Looking Paradigm (PLP), findings reveal that 10-, 14- and 18-month olds categorize goals (and sometimes sources) across motion events, suggesting that these concepts are abstract; not only do infants categorize specific goal paths, "in" and "on" (Casasola, 2008), but they also form the broader category of "goal". In addition to establishing two different measures of categorization, an Intermodal PLP (IPLP) directly measured language development in relation to categorization. The findings suggest that 10- and 14-month-olds do not show evidence of comprehending "to" and "from" in reference to motion events; thus abstract goal and source concepts may be available prior to acquisition of these terms. Second, another study revealed that the goal/source asymmetry of 12- and 16-month olds is influenced by the intentionality of the moving figure (duck vs. leaf) but not by the causal nature of the source (a *cannon* moves the duck) or the physical salience of the source's appearance (bigger, more colorful). Third, the persistence of a goal bias for causal sources was found for children's and adults' language (Lakusta et al., in press). Fourth, current data analyses are directly comparing acquisition of goal/source path terms (measured by IPLP) with goal and source categorization (measured by PLP); the results were presented at CDS (2015). A new study will explore goal and source categorization for non-intentional events with infants 10- to 18-months.

Broader Impact Contributions: The research has impacted society and students, including several students from underrepresented groups. Society: The findings have contributed to understanding language and cognitive development and to theories explaining their development. The methods and findings established under the prior award have been presented to Dr. Gerry Costa, the director of the Center for Autism and Child Mental Health at MSU, and a collaboration is underway to use some of the PI's methodologies (IPLP) to assess children in Early Head Start programs as well as those who have childhood disorders (such as autism). The PI has been made an affiliate faculty at this Center. The research has been disseminated to the general public to educate them about early language/cognitive development; the findings were presented at a local mother's market for two years, at a Lakusta Lab open house in the Spring 2015, and at the MSU Student Research Symposium (2013, 2014, & 2015) which is attended by students, as well as faculty, parents of students, and administration. Further, the PI presented the findings at Ramapo College's Colloquium Series (2013). Student impact: 12 undergraduates and 5 graduate students (all women; S. DiFabrizio, K. Wiseman, D. Spinelli, J. Ardis and O. Garcia) have been supported under this award, and, notably, ~25% of the students are from an underrepresented group (Hispanic); this includes 3 undergraduates (Hernandez, Mendez, Guillermina Ponton) and 1 graduate student (Garcia). All student RAs committed at least one full year to the research and have participated in lab meetings, subject recruitment and testing, data coding, analysis, interpretation, and dissemination (see the "References Cited' document for a complete list with student co-authors noted). Every student presented research at the MSU Research Symposium. Three students traveled to BUCLD (2012) to assist in the PI presentations, 1 student traveled to SRCD (2013) to present

her Master's thesis, 2 students traveled internationally to ICIS (2014) to assist with presentations, 2 students traveled to SRCD (2015), and 2 students attended Datavue and Databrary workshops at NYU (2013) and SRCD (2015). Several undergraduates have pursued further graduate training, including DiFabrizio who is currently pursuing her Ph.D. at CUNY Graduate Center with Dr. Virginia Valian.

Relation of Completed Work to Proposed Work: The completed work explored the breadth and abstractness of goal/source concepts in 10-, 14-, and 18-month-old infants and the relation to language development. The proposed work is an innovative extension of this prior research in that it considers another set of fundamental elements in the encoding of *support*. Similar to goal paths, these representations are also broad and complex, as recently suggested by work in linguistics and cognitive science. Their breadth and complexity raise the question of how infants categorize such events and how children acquire the complex semantic space. Just as the completed work examined the influence of language on goal and source concepts in infants, the proposed work focuses on the influence of language on concepts, but also examines how experience interacting with objects may influence pre-verbal conceptualization of spatial relations. The proposed work will build on methodologies established in the prior work (PLP and IPLP) as well as bring new methodologies to bear on the questions of interest, such as examining spontaneous and elicited production of 1.5-4.5 year olds and examining parent input and experience with objects. Lastly, the proposed work aims to continue the involvement of undergraduate and graduate MSU students (including underrepresented groups), but, importantly, it expands student research opportunities by having two students each year conduct summer research in the Co-PI's lab at JHU. MSU students will collaborate with the Co-PI and JHU students, to increase their skill set, thus offering all individuals a unique opportunity for diverse collaboration (see RUI statement).

Results from Prior NSF MRI Support

2009-2014. "MRI: Acquisition of a High Performance Computer Cluster Supporting Computational Science Research and Learning", CNS-0922644, $190,010, PI Stefan Robila

The funding for this project was used for the acquisition of a high performance computing environment (computer cluster). The high quality of the equipment ensured that the cluster (copou.csam.montclair.edu, or "copou") has continued to operate without any failure since its installation in April 2010. The main cluster was supported by two additional workstations and one Tesla GPU system, all constituting state of the art equipment with unique features for the MSU campus. However, the equipment is now obsolete and will be used mostly for instructional purposes, supporting undergraduate and graduate programs. The equipment was used by six researchers spanning Computer Science, Linguistics, Mathematics and Chemistry as well as by many graduate and undergraduate students [26, 27, 30, 31]. The cluster allowed the PI to redesign and offer courses and independent studies in parallel and distributed computing as well as support courses in Computational Chemistry and Linguistics.

The lead project resulted in novel HPC approaches to processing hyperspectral data, speeding up data unmixing and spectral separation and promises to add to the global research on material identification and classification. We designed and tested a parallel best band selection algorithm (PBBS) using MPI and multithreading. Given n-dimensional spectra $\mathbf{x}$ and $\mathbf{y}$ with the components $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$, metrics for the 'distance' between $\mathbf{x}$ and $\mathbf{y}$ as a function on the spectra' components are designed [105]:

$$d(\mathbf{x},\mathbf{y}) = f(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n) \quad (5)$$

Measures of separability are often found in the literature. One such popular distance is the spectral angle [106]. Since many of the vector components are strongly related to each other, methods that select

only those bands that increase (or decrease) the separability are sought. Given two spectra **x** and **y** with values over a set of spectra bands **B**, a spectral distance β the goal of band screening is to find the subset of bands $\mathbf{B}_1$ such that:

$$(6)\ ),,(\min(),,(_{11}ByxByx_{BB}\beta\beta_{\subseteq}=$$

where by $\beta(\mathbf{x}, \mathbf{y}, \mathbf{B}_s)$ we refer to the value of the distance measure β computed between the two vectors over the subset $\mathbf{B}_s$. The goal in band screening is to find the subset of bands that maximize the distance between the two spectra [105]. While finding the optimum subset is always possible through exhaustive search, the complexity of such operation is prohibitive. Given a hyperspectral image of n bands, we get 2_n possible combinations. Four experiments were performed in order to test whether the PBBS algorithm within an HPC environment provides a significant improvement over traditional sequential methods, as well as to examine the behavior of the algorithm for different inputs. A full description of the experiments is available in [31] and shows significant speedup.

The auxiliary projects also add to the respective discipline's research progress. For example, Feldman's project on morphological tagging dealt with language technology for archaic language varieties, historic language, dialects, etc. Using the cluster, and a new Czech lexus, Feldman generated a lexus for Old Czech thus providing a computationally effective solution to a considerably complex problem [82].

The project provided learning experience to both graduate and undergraduate students. During 2012-2013, two REU participants (Kordecki and Gallo), working with Robila continued to improve on the computational performance of the PBBS algorithm. Finally, the PI involved students in HPC programming through independent studies and MS projects [27, 30, 31]. In a separate line of research the cluster and workstations supported the development of a decision support system for IT staff with the aim of energy conservation [107-109].

2010-2011: "MRI-R2: Acquisition of Scientific Computing Capacity", DMS-0959461 $129,372, PI: A. David Trubatch. Co-PI Chunguang Du

The grant supported the acquisition of the HPC cluster (kruskal.montclair.edu or 'kruskal') which has been used by the PIs and additional researchers at Montclair State University for a diverse collection of investigations.

Nonlinear Waves and lattice Dynamics PI Trubatch, completed a thorough and systematic simulation [110] of lattices derived from the Kotrweg-de Vries equation (KdV). While solitons (coherent nonlinear waves) were first discovered in simulation of KdV lattices, the original goal of these simulations was to shed light on recurrence that had been observed in lattice-dynamics simulations, in apparent contravention of the hypotheses of statistical mechanics. The systematic simulations definitively show that, despite sixty years of investigations of recurrence, solitons and related phenomena, generally accepted, if non-rigorous, hypotheses about the mechanism of the observed recurrences are inconsistent with the results of a systematic study of the dynamics. In addition to a master's degree student who worked on the KdV simulations, two additional masters-degree theses have been completed with the use of kruskal under the supervision of Trubatch. Two of these masters-degree students were first-generation college students who subsequently enrolled in doctoral programs.

Bioinformatics Du carried out intensive DNA sequence analysis with kruskal for his NSF funded project "A sequence-indexed Reverse Genetics Resource for Maize: a Set of Lines with Single Ds-GFP Insertions Spread Throughout the Genome". The project has led to the development of a useful, publicly available, bioinformatic tool, InsertionMapper [54] that allows investigators to assign sequences from the huge output of NGS machines to individual lines carrying single gene knockouts. One hundred and sixty lines carrying a transpositionally active, marked Dsg element were generated and mapped to the genome. All 20 maize chromosome arms are represented among them and the distribution of platforms is random

across the genome. Seed for 60 platforms have been sent to the Maize Genetics Stock Center and sufficient seed for the remaining 100 will be produced in the summer of 2016. Over 1000 Dsg transpositions were produced from every platform. Around 90% have been confirmed as concordant germinal events by testcrosses, and mapped relative to the original platform. All 3230 of the sequence-indexed single-gene knockouts that have been identified with the InsertionMapper are listed on the project's website. 1943 lines for which we had enough seed are available from the Co-op and an additional 3000 will be available in the fall of 2016. All information on sequence-indexed sites, including matching transposant lines, is shared online in [111]. Using kruskal, MSU undergraduates have gained 'hands-on' experience in a bioinformatics course developed by Du (60 students) as well as in independent research projects (10 students). Students who are members of groups underrepresented in STEM (approximately 30% of the students) have participated as project leaders, postdoctoral investigators, and undergraduate students, generating most of the gene annotation in our current database. Ten undergraduates have moved on to graduate work. Every year, participating students have presented their results at the Annual MSU Undergraduate Research Symposium, the Metropolitan Association of College and University Biologists Conference, as well Maize Genetics Conferences. This project has created a sequence-indexed reverse genetics resource, considered essential by the maize genetics community.

Du also used kruskal to develop the Helitron transposable element discovery tool, HelitronScanner [55]. Helitrons are unusual rolling-circle eukaryotic transposons that often capture gene sequences, making them of considerable evolutionary importance. Unlike other DNA transposons, Helitrons do not end in inverted repeats nor do they create target site duplications, so they are particularly challenging to identify. HelitronScanner is a major improvement over previous identification programs based on DNA sequence or structure and it has been used to identify 108,202 Helitrons from a wide range of plant and animal genomes in a highly automated manner. HelitronScanner has great promise as tool in future comparative and evolutionary studies of this major transposon superfamily.

Stochastic Modeling of Disease Forgoston and collaborators have used kruskal to simulate the dynamics of stochastic models of infectious disease spread in populations as part of the NSF-funded investigation "Understanding the Dynamics of Stochastic Disease Spread in Metapopulations". These investigations have, to date resulted in three publications [112-114]. In addition, two students have used kruskal in completion of Master's theses on related investigations of stochastic disease models, one of whom subsequently enrolled in a doctoral program.