Goals and Motivations

As a defining property of human social groups, language is an essential component of any project that attempts to model human behavior. The joint modeling project being undertaken by the Advanced Simulation Technologies Center (ASTC) at the Argonne National Laboratory (ANL) and by the Chicago Language Modeling Lab (CLML)[1] at the University of Chicago aims to integrate a realistic model of language (the language module) into a correspondingly realistic model of settlement systems (ENKIMDU). The ultimate goal of this joint modeling project is to develop new tools for understanding language change and societal change as parts of a larger, interactional system.

Significant initial progress has been made by both the ASTC and CLML groups in developing and integrating a computational model of language into the existing ENKIMDU simulation of ancient Mesopotamian settlement systems[2]. This simulation is based on the Dynamic Information Architecture System (DIAS) framework, and it seeks to realistically model stability and sustainability of ancient Mesopotamian settlement systems by concurrently modeling dynamic factors such as the atmosphere, climate conditions, and spatiotemporal movement of agents as weather, hydrology, crop growth, and human agent spatiotemporal movement and social behaviors, including behaviors associated with response to catastrophes such as drought, famine, plague, and war. However, as the medium of information transfer between agents, language is a crucial element in the model because it affects the responses to these factors and events. This is why our goal has been to develop a stand-alone language module to insert into the existing ENKIMDU simulation.

The ENKIMDU simulation and the language module mutually benefit from each other's capabilities. The ENKIMDU simulation becomes more realistic when a language module is in place to transfer the information that affects the actions of its agents. Effects such as misinformation, lack of communication, rumors, etc. are all dependent upon a realistic information transfer model, and such a model has language at its core (if not as its sole component). While ENKIMDU thus benefits from a realistic language module, efforts at language modeling benefit from the extra-linguistic factors that the ENKIMDU simulation provides. Factors such as prestige gradations, population dynamics and movement, and geographic settlement flux all contribute to language birth, death, and changes to the grammars and lexica of languages. This enriches the understanding of how language and society interact and influence each other. The computational approach that ASTC and CLML utilize in this effort is representative of the growing movement to lend empirical, computational evidence and predictions to existing linguistic and societal research efforts.

[1] CLML website: clml.uchicago.edu
[2] The ENKIMDU simulation framework: oi.uchicago.edu/OI/PROJ/MASS/Mass.htm

ASTC Progress

The ASTC group, headed by John Christiansen, has made significant progress on extensions to the ENKIMDU simulation to support linguistics modeling. In terms of computational performance, the simulation currently utilizes most processing power for landscape processes, such as weather, crop growth, etc. However, the group is changing this situation by adding significant computing power and back-end programming for social processes, such as farming and herding practices, kinship-driven behaviors, trade, etc. As of the writing of this report, 100-year simulations run in 1 to 2 hours on a desktop PC with a 3 GHz CPU. However, the ASTC group is pushing to the envelope of current computational power on several fronts. First, the group is working on implementing simulations of larger geographic areas, such as entire regions instead of single settlements. These will require not only more raw computing power (more processors, more hard disk space, etc.), but also additional programming entities to model new types of dynamic processes, such as inter-urban commerce and migration. Second, the scale of the simulation will also be extended on the micro-scale in order to delve more deeply into finer-scale aspects of the model, facilitating higher-fidelity modeling representations. One important aspect of this is that it will provide finer-grained temporal and social distinctions that will influence the usage of the language module implemented by CLML. For example, it will allow modeling in terms of minutes rather than hours, and this would shed light on conversation dynamics in a more realistic timeframe. Third, ASTC's enhanced ENKIMDU simulation will shortly begin to make use of multiple processors to support the increased computational demands of the expanded model formulation. The growth of the simulation to this extent foreshadows the future need for high-performance computing power, such as that offered by the BG/L "Blue Gene" supercomputer at the ANL Mathematics and Computer Science (MCS) Division.

The language module will also be a factor in increasing the need for a great deal of computing power. The computing power that will be required for just two language agents, one teacher and one learner, will be significant. On top of that, there must be at least 11 two-way processes operating any time two agents engage in linguistic interaction. The perception and articulation component of the module will mutually interact with both the grammar and learning components of the module. The grammar and learning components will mutually interact, and both these components will interact with the part of the language agent that is composed of knowledge and drives. Given two participating agents, this accounts for 10 of the 11 two-way processes, and the final two-way process is the actual transfer of information between the two agents. Although the ASTC's present nanocluster of sixteen 2.4 GHz CPUs can handle the interactions in a small, isolated village with only a few households, handling the interactions on urban or regional scales will require supercomputing power such as that offered by the "Blue Gene" supercomputer.

The ASTC at Argonne will initially integrate the language module developed by the CLML into each simulated social agent, and will also focus on extending ENKIMDU's social models to address spatiotemporally finer-scale interactions to support the generation of more realistic linguistic cliques for the social language agents. These linguistic cliques are situations in which agents interact using language in ways specific to their environment. Two facets of this latter focus are: first, to further define fine-scale agent social behaviors, especially the capability to automatically generate agents' agendas, detailed itineraries, etc.; second, to further define the fine-scale physical environments in which the agents operate. This entails beginning to build spatially explicit models of ancient Mesopotamian urban environments.

In relation to the first aspect, the ASTC has already generated a detailed mapping of different linguistic cliques that will be available to agents in the simulation. These cliques include the neighborhood, the market, the temple, nomad camps, other settlements, etc. Other linguistic cliques prompted by events rather than places will be incorporated into the model, such as chance encounters on the road, traveling in a caravan, military campaigns, pastoral tasks, agricultural and public works tasks, and forays into the wild for fishing or hunting.

In addition to designing ENKIMDU extensions to support modeling linguistic cliques, ASTC has also developed a detailed schema for time allocation according to the type of agent. Major categories of agent that will be modeled separately include, as a minimum: infants, boys, girls, adolescent and adult males, adolescent and adult females, and male and female elders , with further differentiation by social rank (e.g., elite / common / slave), and, as appropriate, profession.

Time allocation is modeled using activity budgets, which allocate not only time, but also the amount of effort and commitment that each agent has the capacity and inclination to endure. These budgets factor in the nominal duration of the activity, its relative priority, urgency, context or causality, frequency of the event, and intervals at which to update each activity's urgency and priority. These activity budgets can be used to automatically generate agent agendas and itineraries, and they drive the agent's actions. One can expect to see systematically different patterns of allocation according to the different social roles that each agent plays.

In further defining the spatial environment in which the agents operate, the ASTC draws on a variety of resources. Notably, it utilizes Oriental Institute (OI) graduate student Tate Paulette's ongoing research on urban textures, as well as the work of the OI CAMEL Lab, which focuses strongly on spatial themes. In addition, the OI/ASTC "SHULGI" JTI project, which seeks to model pedestrian transport in a full social context, is a highly valuable and relevant resource for our project. Finally, the ASTC is able to capitalize on years of spatial and mobility simulation research that has been undertaken at Argonne for various purposes, including military applications. Drawing on all these resources, Argonne is constructing a preprocessor engine to automatically generate representative spatial textures and transportation network topologies for ancient urban areas which includes embedded transportation networks that include not only public thoroughfares but "shortcuts" that may be utilized depending on the social context, such as a social agent knowing the owner of the territory through which he or she hopes to pass.

CLML Progress

The Chicago Language Modeling Lab (CLML), directed by Jason Riggle, has constructed a basic language module that is ready to be incorporated into the ENKIMDU model. The current model is written in Jython, which is a Java implementation of the Python programming language. The model is currently limited to expressing phonology, an aspect of language that has received considerable attention from a computational perspective in linguistics. This is taken to be the grammar of the language. It implements Optimality Theory, a theory of phonology which has been in continuous development since its conception in the early 1990s by Alan Prince, John McCarthy, and Paul Smolensky. This theory conceives of a phonological grammar as a set of constraints that restrict the surface form of a language to a general shape. When the inputs supplied to the model are modified by the constraints that evaluate them, the change can be detected computationally and can be modeled using regular expressions. This allows the model to detect language change among agents, which makes factors that influence language change clearly visible. The language module that the CLML has developed has already been successfully incorporated into DIAS, the framework that underlies ENKIMDU.

The discussion that is currently taking place between the CLML and the ASTC is about the points of interface at which the code for ENKIMDU and the language module needs to be "wired together." The three main points of interface are: production, perception, and the medium for communication.

With regard to production, Argonne's ENKIMDU model will supply information on the context of agents' speech events, for example, when to talk, who to talk to, and what to talk about. The language module will respond with a series of simulated events involving all the components of the language module to determine the grammatical form of what is said, as well as its physical articulation and perception by the agents. In the course of its operation, the language module will be modified, since such modification occurs during a conversation in real life and communicating subjects are constantly unconsciously modifying the way they produce their language.

Perception is a facet of the model in which the social context that ENKIMDU provides will play a large role. Perception of speech will affect efficiency in cooperative tasks, such as bargaining and warnings. Differences in perception will occur in different settings and for different types of speech. For example, ambient marketplace speech must be treated separately from focused negotiation. Nomads present an especially interesting situation for perception, since they may speak a foreign language with which no agent in the settlement is familiar. Nomads, who are somewhat separated from the context of settled life and possibly from the geographic area of the settlement, will cause suboptimal perception if their speech is too different from that of the settlement's agents. Agents will have to respond differently to the nomads than to their normal settlement companion agents. Nomads will also present an interesting stimulus for linguistic change because they will prompt a need for understanding, but will probably have some speech differences. The nomads may also adopt features of the speech communities which they visit.

The third point of interface between the language module and ENKIMDU will be the medium through which speech is conducted. This will affect which agents can receive the signals of other "speaking" agents. However, the "speech" signals of other agents will not be simply spoken speech, but also written communication. Modeling media in this way can later quite easily be extended to address modern media such as text message, cell phone conversation, email, and television. Only some agents will be able to receive the "speech" signals sent by various media, and this part of the model focuses on language, but must rely on ENKIMDU for designating which agents may receive and send such speech signals. This will be accomplished by designating which agents are literate or have the social status required to be able to communicate using the given media.

Conclusion

ASTC has made strong progress and is continuing to work on deepening and enlarging the scope of the ENKIMDU simulation. The CLML has made significant progress on an initial language model that has been implemented in the DIAS framework and will be incorporated into ENKIMDU. Both groups are continuing to collaborate and formulate plans for future growth and research, especially at the interfaces between each group's contributions to the overall simulation. These new contributions will enhance the realism of the model by incorporating elements that have never been included in simulations before, and will facilitate many new discoveries about how settlement systems develop and how language affects that development.