2019　2020

How does the social-communicative environment impact early language acquisition? An ecological perspective

Tsuji, Sho

2,200,000

SCALa

Tseltal　SCALa

How infants learn language so well is still an unresolved mystery. Day-long audio recordings offer a glimpse into infants' environments. We developed an analysis framework for such data, which will allow a better understanding of how environmental cues interact for successful language acquisition.

The present project aimed to create an infrastructure to measure and analyze infants social-communicative environment across cultures. As a first step, we developed a framework to extract such data from the input in a unified way and to link it to known learning algorithms. As a second step, we proposed to collect infant data and apply our framework. Although we were able to do some pilot data collection, due to the pandemic we were not able to collect a full dataset in Japan. Instead, we moved on to the cross-linguistic part of the project and applied our framework to Tseltal, a language spoken spoken in a smallscale, subsistence farming community in Mexico. These data reveal a high quantity of input that could be used to extract language structure, but less input that would permit the learner to link this structure to events or entities in the surrounding world. Together, this approach will allow us to make precise recommendations for future large-scale empirical research.

day-long recordings  language acquisition  supervised learning  unsupervised learning  language structure  social constructivism

How infants are capable of learning language with a speed and efficiency unrivaled even by state-of-the-art AI systems is still a mystery. Their use of social-communicative cues has been suggested as crucial, but little is known about the actual composition of infants' social-communicative environment. The present project therefore set out to measures and analyzes infants' social-communicative input in an ecologically valid setting and to asses its impact on key language outcomes.

Language is one of humans' core cognitive abilities, and how infants across cultures are able to learn it with such unrivaled speed and efficiency remains a key question. One potentially key explanation for infants' superior performance is their use of social-communicative cues. Indeed, past research has demonstrated that such cues augment language learning in the lab. The mechanisms through which the social-communicative environment supports early language learning are, however, still poorly understood. The present project aimed to elucidate these mechanisms by asking three key scientific questions.
**Q1**: What is the content of infants' social-communicative environments?
**Q2:** How does this content relate to infants' language outcomes?
**Q3**: How does this content differ between cultures?

### 3.1 Original plan
The original plan related to the above three questions was as follows.
As to **Q1**, we planned to day-long audio recordings with accompanying hour-long head-mounted camera recordings with wearable devices in Japanese infants' homes. We further planned to extract data from these recording with a combination of automatic and manual annotations. We then planned to devide a coding scheme in order to extract information relevant for subsequent classification such as *speaker identity* (e.g., mother, father, sibling), *content* (e.g., referential domain, such as visible, audible; speech act, such as statement, questions), and *context* (e.g., objects or people in visual field).
As to **Q2**, we planned to assess the same infants' speech sound discrimination and word recognition in the lab Finally, for **Q3**, we planned to extend the analysis scheme we developed to data collected in other languages and cultures.

### 3.2 Changes due to the pandemic
In order to carefully prepare the longitudinal study, Year 1 was dedicated to planning and piloting. Once we were ready to start the actual data collection, it became impossible to invite caregivers and infants to the lab. Note that, although the recordings themselves would have been conducted in infants' homes, we would still have needed infants to come to the lab regularly for the experimental assessment of their language outcomes.
We therefore changed the research plan to what was realistically possible during the pandemic situation, and decided to
**(C1)** Develop a more sophisticated framework for coding socio-communicative information in the input in a cross-cultural context and link it to computational algorithms. This contributes to **Q1** and will facilitate coding once we are able to collect data.
**(C2)** Start coding data in other languages and cultures. As the pandemic progressed, we decided to not wait for the collection of our own dataset, but start to apply our coding scheme to data from other languages and to thus make progress on **Q3**.

### C1: Framework for coding socio-communicative information and link it to computational algorithms
Together with my collaborators Dr. Alejandrina Cristia (expert on cross-linguistic assessment of language input) and Dr. Emmanuel Dupoux (expert on computational approaches to studying language learning), I developed SCALa (Socio-Computational Architecture of Language Acquisition). It situates an ideal learner within an environment, which includes the physical world as well as one or multiple interaction partners that provide social-communicative input to the infant. As illustrated in Figure 1, the learner is imbued with two sets of components: (1) probabilistic models and associated learning algorithms, (2) data

preprocessing components. In the context of a social learner embedded in the real world, the data preprocessing problem is barely specified and complex. The learner requires advanced operations like referential selection and communicative event classification to process input adequately. This is addressed by the classification scheme in Figure 2. Here, we try to systematically enumerate the different cases that could arise depending on the referential and communicative aspects of a simple interaction event of this sort, and link each case to the learning algorithms that would be most appropriate in this particular situation. The results of this projects are now accepted for publication in the journal *Cognition*.
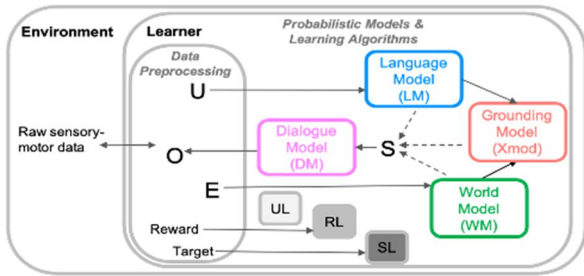


Figure 1. The learner's internal probabilistic models and learning algorithms. U: utterances, E: events, O: outputs

## C2: Applying SCALa to a Tseltal dataset

Lacking own data collected in Japanese, I started collaborating with Dr. Marisa Casillas, who has assembled and annotated a dataset of day-long audio recordings in Tseltal, a language spoken spoken in a smallscale, subsistence farming community in the highlands of Chiapas (Southern Mexico). Based on the society's many characteristics that diverge from a Japanese urban population, it will provide a relevant comparison for the cross-cultural comparison. I developed a scheme for native Tseltal speakers to code caregiver utterances into the scheme developed in SCALa (Fig.2). One example of coding instructions is given in Figure 3. By coding utterances based on our classification scheme, we are able to assess what kind of information is frequent or infrequent in the input. Preliminary results are presented in Figure 4. In Figure 4A, we can see that utterances referring to the sensory domain, thus that refer to something that can be sensed in the moment of the utterance (see example in Fig. 3), are the most frequent. This would mean that they can be not only leveraged by LM (language modeling), but also Xmod (Cross-modal learning) and would allow to learn about associations between an utterance and events in the world. We can also see (Fig. 4C) that instances where feedback is provided are relatively rare. Feedback would allow an utterance to be processed by RF (reinforcement learning), which can be a more powerful learning algorithm compared to unsupervised instances. By comparing the prevalence of different kinds of utterance types across children and cultures, we can start understanding the role of the socio-communicative environment in language acquisition.
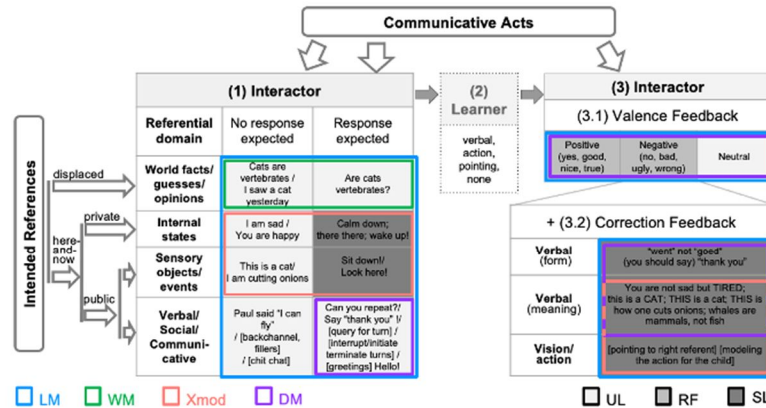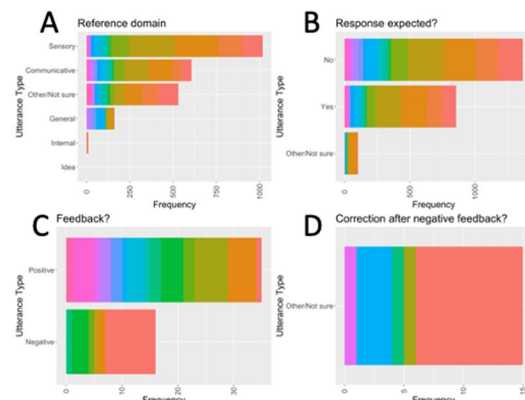


Figure 2. Possible events based on reference and communicative intentions and how they would relate to various probabilistic models and learning algorithms in an ideal learner.



Figure 3. Example coding instruction with examples for the dimension "Intended references"



Figure 4. Preliminary results of SCALa classification in Casillas' Tseltal corpus.

| | | | |
|---|---|---|---|
| | | | |

0

| | | | | |
|---|---|---|---|---|
| | Ecole Normale Superieure | INRIA | EHESS | |
| | University of Chicago | | | |
| | CIESAS | | | |