Consistency and Variability in Language Learning

How do wordless infants develop so quickly into toddlers who can use language to communicate? Every typically developing child acquires the ability to use language. Yet children vary dramatically in both the trajectory of their learning and their eventual attainment, with subsequent consequences for important abilities like reading. No theory of children's language learning explains both the consistency of the process and the variability in its outcomes.

In search of such a synthetic view, my research aims to create data-driven computational models of language learning that link children's social and environmental input to their attainment. Such models make theoretical assumptions explicit, providing a method for cashing out proposals about what learning mechanisms are shared across individuals and how learning outcomes are predicted to vary with differing environmental inputs. Working towards this goal, I create and curate datasets that provide constraints on theory (§1) and then develop models that can exploit these datasets (§2), looking for synergies between data-gathering and model-building. Our models in turn make predictions that can be tested with targeted developmental experiments (§3). Theoretical work is only meaningful if the empirical foundations are solid, however; thus, I also study – and advocate to improve – the reproducibility and replicability of developmental science (§4). The overarching goal of my research is a better understanding of the fundamental principles underlying the consistency and variability of early language learning.

1. Large-scale datasets describing early language development

Creating quantitative theories of early language learning requires datasets that capture the broad arc of early childhood development. My lab's work has been focused on creating datasets that characterize the growth of children's language both observationally and experimentally.

Large-scale vocabulary outcomes. To measure child language from birth to age three, many researchers use the MacArthur-Bates Communicative Development Inventories, reliable and valid parent-report measures of vocabulary size and language complexity. To aggregate data from these instruments, we created Wordbank, an open database of more than 75,000 children's data from 29 languages/dialects (Frank et al., 2017, *J Child Lang*). In our forthcoming book (Frank et al., 2021)[1], we describe consistent patterns of acquisition across languages, creating a baseline set of observations as targets for predictive modeling as described below (§2).

Experimental measures of development using meta-analysis. Experimental measures are often the most sensitive probes for specific phenomena, yet individual experiments are noisy. Statistical meta-analysis counters this issue by averaging across an entire literature (e.g., Lewis et al., 2020, *Cognit*). In the MetaLab Project, we organize data from 25 meta-analyses of early language, allowing the construction of theories that depend on the full breadth of the literature rather than a single experiment (Bergmann et al., 2018, *Child Dev*).

Capturing children's linguistic input. Corpora of children's language input are a critical resource for learning models: recognizing this, we have organized transcript data from the CHILDES archive into a reproducible database (childes-db; Sanchez*, Meylan* et al., 2018, *Behav Res Methods*). Yet the input to language learning is children's visual experience as well: to address this issue, we created the SAYCam dataset, a longitudinal dataset of more than 200 hours of head-mounted camera video for each of three children, the largest publicly available dataset of first-person video (Sullivan et al., under review).

2. Computational theories of language learning and use

In my dissertation work, I developed a computational model of children's word-object mapping in social context (Frank, Goodman, & Tenenbaum, 2009, *Psych Sci*). We modeled word

[1] Representative publications are underlined.

learning as a joint inference at two timescales: about speakers' communicative intentions in the moment (pragmatics) and about the stable, long-term correspondences between words and their meanings (vocabulary). These two timescales are the focus of our subsequent work.

Rational Speech Act (RSA) models of pragmatic reasoning. To better understand the in-the-moment social inferences underlying early word learning, I have collaborated with Noah Goodman to develop what we call "Rational Speech Act" models of pragmatic inference (Frank & Goodman, 2012, *Science*; reviewed in Goodman & Frank, 2016, *Trends Cognit Sci*). The basic assumption of RSA models is that language is a form of rational action: a speaker's meaning can be recovered from an ambiguous utterance by considering different plausible goals for what they would like to communicate to the listener. These models provide a tight quantitative fit to adults' judgments about language use in context; they have also been a generative framework for studying linguistic phenomena including common ground (Bohn et al., 2019, *Proc Cognit Sci*), politeness (Yoon et al., 2020, *Open Mind*), vagueness, hyperbole, generics, and others.

Predictive models of vocabulary. Understanding patterns of variability and consistency requires models that predict at the level of an entire vocabulary, rather than a single situation. We initially developed approaches to this problem to predict language outcomes for a single child (Roy et al., 2015, *PNAS*). We have since generalized this method by leveraging Wordbank and childes-db (§1), fitting predictive models of average developmental patterns across 35,000 children learning 10 different languages (Braginsky et al., 2019, *Open Mind*, see also: Wordbank book). This analysis revealed striking consistency in the information sources used by learners acquiring different languages, providing a proof of principle for using large-scale data to inform computational theories. Our most recent work uses more sophisticated semantic network and syntactic prediction models to compare the fit of different computational theories to cross-linguistic data (Fourtassi et al., 2020, *Cognit Sci*; Portelance et al., 2020, *Proc Cognit Sci*).

3. Testing theory through targeted developmental experiments

My collaborators and I have worked to hone existing methods like eye-tracking and develop new ones (e.g. tablet experiments for children; Frank et al., 2016, *J Cog Dev*) to provide precise measurements of children's learning, allowing for tests of predictions from our theories.

Pragmatics and implicature. Much of our experimental effort has been devoted to testing predictions of RSA models (reviewed in Bohn & Frank, 2019, *Ann Rev Dev Psych*). For example, the RSA viewpoint predicts that pragmatic inferences are easy for children – contra prior work that had shown long developmental delays. In a series of experiments, we confirmed that the inferences themselves are indeed easy and can be performed by two- and three-year-olds when the linguistic contexts are simplified sufficiently (Stiller, Goodman, & Frank, 2015, *Lang Learn Dev*; Horowitz et al., 2018, *Child Dev*; Yoon & Frank, 2019, *J Exp Child Psych*). Further, children can use these implicatures to learn the meanings of words (Frank & Goodman, 2014, *Cognit Psych*) and even the structure of concepts (Horowitz & Frank, 2016, *Child Dev*).

Social cues and active learning. The broader pragmatic viewpoint suggests that social information from speakers' gestures and eye-movements should be extremely important for word learning. Indeed, from an early age, social information determines what children learn in simple mapping experiments (Yurosky & Frank, 2017, *Dev Sci*). Yet these cues are not always available (Frank et al., 2013, *Lang Learn Dev*); thus, children actively seek out information as well (Hembacher, deMayo, & Frank, in press, *Child Dev*; MacDonald et al., 2020, *J Exp Psych Gen*).

Function word learning. Most work on word learning has focused on simple nouns like "ball" and "dog." To make a more challenging test of our models, we have tried to understand the learning dynamics of function words like "no," "the," and "or," which cannot be interpreted as referring in the physical context. These words are far more common than nouns, but more data may be needed for children to generalize appropriately (Meylan et al., 2017, *Psych Sci*). Supporting the broader pragmatic view, we find that context is critical for inferring the meaning of "no" (Nordmeyer & Frank, 2013, *J Mem Lang*; Nordmeyer & Frank, 2018, *Lang Learn Dev*).

4. The critical role of reproducibility and replicability

Empirical observations only constrain theory to the extent that they are replicable, yet an initial study suggested that many findings are difficult to replicate (Open Science Collaboration, 2015). Though this study had significant limitations, my role in it convinced me of the critical importance of meta-scientific research in revealing practices that support robust, cumulative science. Subsequent follow-ups support the role of sound statistical and sampling procedures in contributing to robustness, moving the discussion away from factors like experimenter expertise (Mathur et al., in press, *AMPPS*) and towards transparency, openness, and computational reproducibility (Klein et al., 2018, *Collabra*; Hardwicke et al., 2018, *Royal Soc Open Sci*).

For developmentalists, these issues are especially distressing because we build our theories around a relatively small set of experimental findings, often with limited sample sizes. Addressing this issue, I founded the ManyBabies Consortium, a collaborative replication network for developmental psychology (Frank et al., 2017, *Infancy*). We conduct high-quality, multi-site replication studies designed to lead to new theoretical and methodological advancements. ManyBabies 1, a 67-lab replication of the infant-directed speech preference, provides a proof of concept for this model and shows the consistency of this phenomenon across 17 countries (ManyBabies Consortium, 2020, *AMPPS*). This work has led to numerous spinoffs (e.g., with bilinguals; Byers-Heinlein et al., in press, *AMPPS*). ManyBabies projects 2 – 5 are already under way, showing the robustness of this model of collaborative science.

Conclusions

In service of understanding language learning, my work aims to close the gap between data, theory, and experiment. The promise of this effort is the development of new ways of understanding children's language learning that shed light on the uniquely human consistency of language as well as its variability across individuals and linguistic communities.

Contributions to Teaching and Service

I have taken a wide variety of service roles, focusing on open science, diversity and inclusion, and the importance of interdisciplinary cognitive science. At Stanford, I direct the Symbolic Systems program (2020-present). I have been Deputy Director, Acting Airector, and most recently, Director (2020-present) of the Center for the Study of Language and Information and have co-led the center's NSF-funded summer program (2014-present). I have served on the open science and diversity committees of the Psychology Department. I also served as Chair of the Cognitive Science Society (2018- 2019) and served on the founding committee of the Society for the Improvement of Psychological Science. I have been an associate editor at Cognition and am currently special section editor for Registered Reports at Child Development.

In my undergraduate teaching, I have focused on (1) making a developmental perspective accessible to large groups of students in Intro to Developmental Science (Psych 60, taught seven times 2011-2016) and (2) integrating this material with an interdisciplinary perspective in Behavior, Health, and Development (Human Biology 3B, taught in 2019 and 2020). In my graduate teaching I have focused on sharing the skills and concepts that underly reproducible and replicable research in Experimental Methods (Psych 254/251, taught nine times from 2012 – 2020). It has led to numerous published student projects. Its philosophy of teaching replication is described in Frank & Saxe (2012, *Persp Psych Sci*); student replication projects from class were contributed to a class meta-scientific project (Hawkins*, Smith* et al., 2018, *AMPPS*).