# Language input and outcome variation as a test of theory plausibility: The case of early phonological acquisition

Alejandrina Cristia

# Language input and outcome variation as a test of theory plausibility: The case of early phonological acquisition

Alejandrina Cristia[1]

[1] Laboratoire de Sciences Cognitives et Psycholinguistique, Dept d'Etudes Cognitives, ENS, PSL University, EHESS, CNRS

## Author Note

Correspondence concerning this article should be addressed to Alejandrina Cristia, 29, rue d'Ulm, 75005, Paris, France. E-mail: alecristia@gmail.com

Abstract

There is wide individual, social, and cultural variation in experiences afforded to young children, yet current evidence suggests there is little variation in phonological outcomes in the first year of life. This paper provides a classification of phonological acquisition theories, revealing that few of them predict no variation in phonological acquisition outcomes, and thus are plausible in view of observed patterns: Only theories with strong priors and informational filters, and where phonological acquisition does not depend on lexical development, are compatible with great variation in early language experiences resulting in minimal or no outcome variation. The approach is then extended to consider proposals contemplating acquisition of other linguistic levels, including joint learning frameworks, and testable predictions are drawn for the acquisition of morphosyntax and vocabulary.

*Keywords:* input; phonological acquisition; individual variation; socioeconomic status; cultural variation

Word count: 8,600 (excluding abstract, appendix, and references)

# Language input and outcome variation as a test of theory plausibility: The case of early phonological acquisition

Human infants typically grow up to be competent users of their ambient language(s), including its phonology (i.e., the sound system). Before the end of the first year, infants' perception begins to attune to the native language(s), with maintenance or increases of discriminability for contrasts that are relevant, and a reduction of discriminability for contrasts that are not (Tsuji & Cristia, 2014).[1] They also show sensitivity for sound sequences that are frequent and legal, as opposed to infrequent or illegal, in the native language (Jusczyk, Luce, & Charles-Luce, 1994). Evidence suggests that perceptual development is partially governed by maturational constraints (Werker & Hensch, 2015; see also Gonzalez-Gomez & Nazzi, 2012). Over the first year of life, infants' production gains in complexity, from the entirety of early vocalizations being composed of non-canonical syllables, to a sizable minority containing adult-like consonantal transitions (Oller, Eilers, Steffens, Lynch, & Urbano, 1994; Petitto, Holowka, Sergio, Levy, & Ostry, 2004). Numerous theories have been put forward to capture how infants learn their language's phonology in the first year of life, including some with a domain-specific, general-cognitive, and social slant (respective examples for phonology are Hale & Reiss, 2003; Jusczyk, 1993; Kuhl, 2007). Here, I provide a meta-theoretical framework for evaluating data requirements across theories, including what counts as "input" to the language acquisition process and how much of it is necessary to achieve some acquisition state. As will be explained below, since different infants receive widely diverse amounts of language input of different types, theories diverge in their predictions of the level of variability that should be observed in learning outcomes.

[1]The majority of citations and explanations here will make reference to "vocalizations" and "speech", due to the fact that the developmental literature, particularly in the first year of life that is the focus of the present paper, has mainly studied oral language. As a result, we do not know whether phonological acquisition in the visual or tactile modalities follows similar or distinct patterns.

**1.1 Variability in early language experiences.** There are marked differences in the quantity of input directly addressed to children growing up in different socioeconomic groups and cultures (Hart & Risley, 1995; Ochs & Schieffelin, 2001). For instance, in Hart and Risley (1995)'s study, by three years of age, children whose parents were in the highest socio-economic status (SES) group had received three times the child-directed input that children whose parents were on the lowest status group had (28.5 compared to 8 million words, respectively). As for cultural variation, a systematic review suggests there is at least a 6:1 ratio (estimated at 30 to 5 million words; Cristia, 2019).

Additionally, there is no evidence at present for compensatory behaviors that would ensure that experiences provided to a child increase in quality when they decrease in quantity. To illustrate this, let us take word learning as an example. By most theoretical accounts, higher quality input for word learning will include hearing the word form in a diverse set of sentence contexts and having contextual support (e.g., pointing) for words refering to objects in the world. Moreover, it is not the case that children who get less input (lower quantity) hear words in more diverse sentential contexts and with more contextual support (higher quality). A recent review on language experiences as a function of socioeconomic status concludes that American children from disadvantaged households get less speech addressed to them, this speech is less diverse in terms of grammatical structures, and their caregivers use less pointing (Pace, Luo, Hirsh-Pasek, & Golinkoff, 2017). Thus, even considering several quality dimensions, there is no compensation for the lower input quantity, at least in these samples (see Appendix A.1 for further discussion).

**1.2 Variation in phonological outcomes.** The previous section refered to evidence for striking variation in terms of early linguistic experiences: Some children get markedly more input than others, and we cannot assume this is compensated by this smaller input being substantially higher in quality. Therefore, a key question is: What is the extent of variation in phonological acquisition outcomes? Although more work is needed (described

in Section 5), current results suggest that it is not large. For perception, I am only aware of one study: Melvin et al. (2017) found no effect of SES (a proxy for input quantity, given Hart & Risley, 1995's results) on the decline of attention to a non-native sound contrast.

There are more studies on the production side. Let us start with vocalization quantity. Lenneberg (1969) documents little difference in how much infants vocalized in a comparison between infants whose parents were hearing or deaf (in which case, children presumably hear less speech). In contrast, Kim Oller and colleagues found clear differences in vocalization quantities when comparing infants from mid versus low (or very low) SES (Oller, Eilers, Basinger, Steffens, & Urbano, 1995; Oller et al., 1994). However, there are methodological differences between the studies: The former was based on spontaneous vocalizations detected in very long, home-based samples when infants were aged 0-3 months, whereas the latter built on laboratory recordings where caregivers attempted to get infants aged 4 to 18 months to vocalize. Additionally, overall volubility may not be as indicative of phonological development as qualitative aspects of production. When looking at the proportion of vocalizations that were linguistic and/or advanced, as well as the age of babbling onset, Oller and colleagues conclude there are no sizable SES effects (Eilers et al., 1993; Oller et al., 1995, 1994). The lack of relationship between SES and babbling onset is replicated in McGillion et al. (2017). One large-scale study on daylong recordings analyzed with fully automatized software suggests that both vocalization quantity and the proportion of vocalizations that are linguistic differ as a function of SES (Warlaumont, Richards, Gilkerson, & Oller, 2014). However, it is unclear to what extent even the latter reflects differences in phonological abilities and knowledge.

Although I will suggest which further data that would be desirable in Section 5.1, the conclusion based on current evidence is that phonological outcomes are relatively stable, and thus phonological acquisition and/or learning processes are robust to data quantity variation.

**1.3. Summary.** The main aim of this paper is to discuss how variation in input and related variation in outcomes can help us adjudicate between theories: Theories that predict input-outcome relationships that match those that are observed are more plausible than those making incorrect predictions. To assess this, in Section 2, I provide a meta-theoretical framework that allows easy description of different classes of theories and their features. In Section 3, I apply this framework to theories of phonological acquisition, finding that only a small subset of theories seem compatible with our observations of large input variation and small outcome variation in phonology. Section 4 draws broader generalizations on what theoretical features determine how data-hungry or data-lean a theory is, taking into account further phonological development as well as other linguistic levels. Finally, further empirical work is recommended in Section 5.

## 2. A meta-theoretical framework for theory description

In order to decide how data-hungry or data-lean a theory is, compared to other theories, it is useful to have a framework that allows their parallel description. The theoretical features in the framework I propose are:

- Filters: These are innate perceptual and/or attentional filters that allow certain aspects of the environment to enter the child's processing space, and direct this information to a given procedure. In the present framework, these filters could be specific to language, general to whatever modality the child's language travels in, or even more general. They are defined purely by their function: Gating or seeking information in the environment, and directing it to the right procedure. Theories often specify differential processing for different learning procedures and outputs; for instance, one may propose that the child attends to all spoken input, but a filter submits only some aspects of the speech signal to certain procedures (e.g., segmentation into recombinable units to learn words) and other aspects to substantially different procedures (e.g., segmentation into

talkers to learn about conversational structure).

- Priors: These are innate expectations, biases, and knowledge. Like the filters, they need not be domain-specific or even specific to humans; one example is the expectation that speakers may minimize their effort, hypoarticulating words when they are highly predictable in the context, and hyperarticulating them in low-predictability contexts. It simplifies exposition to assume all priors are helpful or neutral; that is, they never mislead the learner.
- Procedures: These are innately specified operations that take into account the input (and if available, the priors) to lead to some learning outputs. They can also be human/domain-specific, or not.
- Outputs: The output of the learning procedure; what is learned or honed or acquired or developed. In some theories, the outputs are exclusively linguistic knowledge, such as an abstract phonological inventory. For others, the outputs are a long-term storage of multimodal exemplars experienced over the course of learning. Yet others would rather describe the consequence of learning as "specialized neural pathways", or "honed language processing skills". Some hold the acquisition process must be completed by a certain age, whereas others believe language skills continue to evolve throughout one's lifetime. But all theories assume that something changes in the mind/brain of a child who has acquired a given language by a given point in time: If one takes a set of identical twins and separates them from birth, twin A, exposed to language A, will eventually differ from twin B, exposed to language B, when faced with one and the same stimulus (both are tested with the contrast between /r/ and /l/). The *outputs* component attempts to capture whatever it is that makes twin A and twin B different at this point of development.

When considering data quantity requirements, we must pay attention both to the initial state, namely specified filters, priors, and procedures, all of which are innate; and aspects of the final state, specifically, the learning outputs, which include any changes to

filters, priors, and procedures. Considering the acquisition process holistically allows us to clearly express that if a theory assumes that filters are changed with age and experience, then the learning outputs must include "acquiring that filter" and the procedures must include a procedure for how the filter is altered (including what information is processed to do so). Similarly, if a theory proposes that children learn a procedure, then the initial state must include procedures to learn that procedure on the basis of experiences and priors; and the outputs must include the procedures that are learned. The same goes for priors: A theory that assumes that the child is natively endowed with no innate priors but learns those priors through experiences must specify learning procedures that extract priors from experience, and thus the learning outputs would include changes to the priors.

## 3. An overview of phonological acquisition theories and their data quantity requirements

To help readers focus on the question at hand (how input-outcome relationships can constrain theory plausibility), I discuss theories globally, in groups or classes, rather than in individual detail. I hope that, in this way, this paper does not come off as an attack on individual theorists, or a proposal that is only relevant to currently existing theories. Speaking about classes of theories should make it easier to see how this approach could be used to evaluate even theories that do not yet exist because they will be proposed in the future.

**3.1 The "innate language networks" class.** The *innate language networks* (ILN) class is based on proposals by Dehaene-Lambertz and Gliga (2004), Hale and Reiss (2003), and Hayes (2004) among others. In general, proposals within this class assume that infants pay preferential attention to linguistic information. This is captured in the proposed meta-theoretical framework as **filters**: Perceptual and/or attentional filters allow/force only certain aspects of the environment to enter the child's processing space. For example,

Dehaene-Lambertz and Gliga (2004) state that infants show a preference for speech over non-speech; this perceptual filter allows/forces infants to not process non-speech information (at least with certain brain networks), which may have been detrimental for learning as it would add noise or irrelevant information. Within speech information, infants attend to both talker and phonemic information, but these two types of information are treated by two different neural, and presumably computational, mechanisms. This example illustrates well that filters do not mean simply ignoring certain types of information altogether, but rather treating certain kinds of information and not others with certain processing functions. I return to the question of whether this prior can really work in realistic language learning conditions in Section 4.1, but for the purposes of the current exposition, I will assume that a "linguistic filter" allows infants to select relevant information in their environment.

As for **priors**, infants would minimally be endowed with a roadmap of the sounds or contrasts that may be found in any human language, as well as "linking rules" that allow them to detect the abstract sounds/contrasts in their environment. (That is, faced with the auditory experience of a [d], infants' abstract /d/ category is activated.) Some of these theories may attribute more to the infant, such as a list of desiderata or "constraints", whose relative importance needs to be established via experience, and which would help infants learn and generalize the sound sequences in their language.

Setting aside sound sequences, the learning **output** is minimally the phonological inventory, selected from the innately available and universal set of contrasts. To find the relevant subset, one **procedure** could go as follows: Infants tally up the frequency with which they have encountered the different surface forms of sounds, either in running speech or looking only at their long-term lexicon for potential contrasts, and decide which categories are "in" and which are "out" based on their frequency of occurrence.

Proposals within the ILN umbrella are the least data-hungry thanks to their strong informational filters and priors, their low expectations in terms of what children learn, and

the fact that they do not distinguish between child-directed and child-overheard input, thus assuming that learners process all linguistic input.

Specifically, in the version of acquisition whereby infants settle on category presence based on sheer exposure to the running input, one can readily imagine a numeric threshold that is higher than non-native sounds but lower than the amount of exposure provided to the child enjoying the least input. Thus, this version easily accommodates the observation of no outcome differences in spite of large input differences.

A different subclass of ILN proposals assume children will decide on the phonological relevance of a sound or contrast not based on sheer exposure to its phonetic form in running speech, but rather depending on whether the contrast serves to distinguish minimally different words (such as bin versus pin). Less overall input means that children hear fewer word tokens, and they probably also hear fewer word types (MacWhinney, 2000, pp. 111–112). This leads to the prediction that minimal pairs should be harder to identify with less input. Thus, versions of ILN relying on lexical knowledge do predict some variation across children as a function of how much input they hear. As a result, this second subclass does not accommodate the observed pattern of large input variation with no outcome variation.

#### *3.1.1 A note on lexically-driven theories.*

Before proceeding, it is worthwhile to discuss aspects of all theories that rely on lexical acquisition. As mentioned above, infants' phonological perception and production changes over the first year of life, as early as 6 months (e.g., for vowels; Tsuji & Cristia, 2014). As a result, to posit that lexical acquisition is necessary for phonological development, one should find that the former develops even before this point. Although there are some published results showing word recognition in the laboratory by 6 to 9 months of age (e.g., Bergelson & Swingley, 2012 and references therein), early recognition effect is not based on word forms

that are well specified phonologically, since 6-7-month-olds succeed to a similar extent regardless of whether the word is mispronounced or correctly pronounced (Bergelson & Swingley, 2018). Together with the general observation that the early lexicon is extremely sparse, results like this one have led many theorists to posit that infants' phonological development does not await the development of a lexicon (e.g., Jusczyk, 1993; Werker & Tees, 1984). Nonetheless, one could propose learning strategies that do not require minimal pairs or even a lexicon with sufficient phonological neighbors for these to exert a pressure for phonological refinement (e.g., Swingley, 2005). Therefore, in what remains in this paper I will set aside whether lexically-driven theories are plausible at all (given the timeline of acquisition), and assume instead that they are in principle plausible.

There are numerous theories on different aspects of infant and toddler word-learning, and it would be challenging to even cite a few of these without failing to acknowledge the diversity of views on this topic. By and large, however, input plays a crucial role in all theoretical views I am aware of (including proposals that highlight bottom-up statistically driven mechanisms, such as Smith & Yu, 2008; theories that instead highlight biases infants bring to the task, such as Trueswell, Medina, Hafri, & Gleitman, 2013; and others that accentuate the social situation in which word learning may occur, Yurovsky, 2018). This is reasonable because every language's lexicon is different from every other language's lexicon; this is clearly one domain in which infants need to gather information from their environment to learn what concepts and word forms are present in their native language. As a result, even if the precise degree of variation predicted may vary across theories, all theories predict individual, social, and cultural variation such that children who get more language input should meet lexical benchmarks faster. It follows then that any domain of language that crucially depends on the lexicon will likely exhibit individual, social, and cultural variation.

**3.2 The "bottom-up phonology" class.** The *bottom-up phonology* (BuP) class is inspired by the theoretical and empirical results found in Maye, Werker, and Gerken (2002),

and a variety of computational models (see sections 3 and 3.1 of Räsänen, 2012's comprehensive overview o phonological and lexical acquisition models). Kuhl's early Native Language Magnet proposal (Kuhl, 1991) would also belong here (NLM-e, the version of Perceptual Magnet that is socially driven, is discussed in Section 3.4). Finally, approaches in which phonological categories are derived from a proto-lexicon may also be included here (see Fourtassi & Dupoux, 2014 for a computational model and @frank2014weak for experimental results). The latter differ from lexical proposals discussed in Section 3.3 because they only assume that infants learn the distribution of sounds or syllables found in items segmented from the running speech, often called word forms although they may also include part words and (part)word combinations that occur frequently. Thus, they do not assume that infants learn their associated meaning, grammatical class, or syntactic collocations.

Innate **priors** are much weaker for BuP than ILN: Infants distinguish a wide range of sounds even without exposure (similarly to ILN in effects, although the reason is different, e.g., domain-general auditory sensitivies), but there is no assumption that infants are endowed with a native "linking rule" between perceptual-articulatory experience and abstract phonological categories. **Filters** are broader than in ILN: Infants are thought to be attracted to primate vocalizations or gestures.

Since these priors and filters are weaker than those in ILN, this class of theories necessitates modifications to the filter over the course of acquisition, in order to eventually exclude human non-speech and non-human vocalizations from the phonological acquisition workspace. These modifications were not needed in the ILN proposal (the innate filter was already focused on linguistic input). Notice that this example highlights a trade-off whereby a lighter prior leads to heavier **outputs** (which must now include a modifcation of the filters) and **procedures** (which must now contain learning strategies to use input in order to modify the filters).

As for **outputs**, the child is thought to derive some form of speaker-invariant,

context-invariant phonological representation, similarly to ILN proposals. To this end, the **procedures** presumably involve segmentation of the input signal into phone-like units and clustering, with no need for first deriving a vocabulary. In the case of protolexical BuP, this is followed by reanalyses of the protolexicon, which may also require segmentation and clustering, albeit of units larger than the phone.

As a result, proposals within the BuP class have a slightly superior level of data requirement than the non-lexical ILN class, due to their weaker priors and filters. It is unclear where they are placed in comparison to the lexical ILN proposals, because this would require establishing tradeoffs between the increases in data requirements due to weaker priors and filters in BuP compared to the increases due to lexical development in lexical ILN. In any case, it is clear that proposals in the BuP class are more data hungry than non-lexical ILN, and thus less compatible with equal outcomes despite wide input variation than the latter are.

**3.3 The "bottom-up lexicon and phonology" class.** Yet another level up in terms of amount of data required are proposals within the *bottom-up lexicon and phonology* (BuLP) umbrella. There are numerous proposals falling within this class, including the "Word Representation and Phonetic Structure Acquisition" (WRAPSA) model proposed by Peter Jusczyk (e.g., Jusczyk, 1993); the phonetics-phonology proposal by Pierrehumbert (2003); the "Processing Rich Information from Multimensional Interactive Representations" (PRIMIR) model espoused by Werker and colleagues (Curtin, Byers-Heinlein, & Werker, 2011; Werker & Curtin, 2005), and a wide range of computational models (notably, some models dicussed in sections 4 of Räsänen, 2012).

**Priors** are weaker than in ILN but potentially equal to those in BuP, with the exception that presumably infants need to be endowed with some expectation relating sounds and words, given that BuLP assume infants learn words jointly with sounds. **Filters** are comparable to those in BuP and thus theories in this class would also require a filter

modification over the course of acquisition.

In addition, the target knowledge attributed to the child (the **outputs**) is a great deal richer in BuLP theories than in ILN and BuP. To begin with, theoretical proposals in this class tend to attribute to the child graded representations of sound frequency and distributions, not just presence/absence of sounds, which means that more data is required for the child to compute the appropriate distribution. For example, if the child expects sounds' frequency of occurrence to be a single Gaussian distribution, she must now estimate two parameters (mean and standard deviation) rather than following a binary presence/absence criterion. But often, theoreticians under the BuLP umbrella propose that much more is stored by the learner, including contextual effects within the same linguistic signal (e.g., sounds in specific words) and outside it (e.g., sounds as a function of the talker). To a certain extent, sociophonetic representations within BuLP proposals are "cheap" in that newborns are thought to represent speech holistically, conflating linguistic characteristics and talker characteristics. Nonetheless, this does mean there are many more, and more complex, distributions to track than in ILN and BuP.

As their name indicates, bottom-up lexicon and phonology theories typically assume that truly productive, stable, and strong phonological representations can only be derived from the lexicon, and thus these theories must also specify **procedures** for how children learn words. As mentioned above for the lexically-dependent ILN theories, this also entails a greater dependence on data, predicting greater disparities as a function of input quantity.

As a result, BuLP proposals predict at least some degree of outcome variation due to several of their features (weaker filters, weaker priors, procedure dependent on the lexicon), and thus do not match the observed patterns for phonology (great input variation, no outcome variation).

**3.4 "Socially-gated" varieties.** This class of proposals is most strongly inspired by Kuhl (2007), although a range of computational models can be thought to belong to this class as well (though often these are focused on language evolution rather than ontological development; see Kaplan, Oudeyer, & Bergen, 2008 for an overview). Conceptually, socially-gated proposals are variants of the other classes (e.g., ILN-SG, BuP-SG, BuLP-SG), as their sole distinguishing feature is the fact that they propose that children do not learn from all experiences, but only from some that are tagged as social. In other words, one can transform the proposals above into socially gated versions by adding a **filter** allowing an experience to enter the workspace only if the experience is marked as "social". For example, experiences entering the workspace need to have features such as being spoken with Motherese intonation or gestures, and involving direct gaze or touch to the child. Other than this addition, then, the exact same architectures and considerations contemplated for the ILN, BuP, and BuLP classes would be relevant here.

Since social experiences are a subset of all experiences afforded to the child, proposals within this umbrella will apparently require more overall input (i.e., directed + overheard) to reach the same levels of outputs given the same priors and procedures as a non-socially-gated variant. This is because in the SG versions, overheard input is discarded, and does not count.

Provided that they do not rely on the lexicon, SG versions of ILN can still accommodate findings of great input variation and no outcome variation: The threshold used to decide whether a sound is present in the environment needs to be lower than the amount of input provided to children who are talked to the least. In contrast, SG variants of ILN relying on the lexicon as well as SG variants of BuLP both predict variation in outcomes when there is variation in input quantity, and thus neither matches the observed pattern.

**3.5 "Feedback-based" varieties.** This final class follows proposals by Majorano, Vihman, and DePaolis (2014), Gros-Louis, West, Goldstein, and King (2006), Goldstein and Schwade (2008), and Warlaumont et al. (2014). The key feature of this class is that the

learner relies on reinforcement signals provided by those in the environment. Infants trigger the onset of a learning experience when they produce a vocalization (including babbling). Phonological acquisition then occurs as one or more of the following takes place. First, caregivers may differentially reinforce (or selectively encourage) more advanced over less advanced vocalizations. Second, caregivers may provide specific feedback, e.g. imitate the child or recast the child's production into an adult-like phonology, thus providing the child with a sort of "label". Finally, the child's output itself may serve as anchor for perceptual representations, meaning that the child is more likely to notice in the others' productions the sounds and sequences the child herself can already produce (see Appendix A.2).

These differences in procedures are important, but a different framework would be needed to capture them fully (see Tsuji, Dupoux, & Cristia, 2019). Nonetheless, they can be set aside to focus on one common aspect of this class which is crucial in the context of input quantity differences: the fact that the child's learning procedure supposes a cooperative tutor.

I will assume that the general idea that there could be feedback loops via production and tutors' responses is compatible with ILN, BuP, and BuLP definitions for innate **filters**, **priors**, and **learning outputs**, and thus propose to present feedback-based theories as special cases of ILN or BuLP proposals (i.e., ILN-FB, BuP-FB, BuLP-FB). Before moving on, I admit that there are some other differences in architecture and scope. Specifically, proponents of feedback-based phonological acquisition tend to propose more modest learning **outputs** than BuLP, notably without necessarily assuming the maintenance of a lexicon and its effects on the phonology. Nonetheless, for the purposes of the current exposition, the presentation is clearer if again we assume there exist feedback-based and non-feedback-based ILN, BuP, and BuLP proposals.

Given the necessity of awaiting feedback instances, even fewer datapoints in the environment enter the acquisition workspace here. As for variation, since reinforced instances

are necessarily a subset of child-directed interactions, then it must be the case that children who get a great deal of attention from caregivers profit from more learning instances than children who get less attention. In fact, a more complex model, which also considers priors, filters, and procedures in the *caregiver* is needed to fully work out the additional level of variation that should be observed according to feedback-based theoretical variants (see Appendix A.2 for brief discussion). For example, since FB variants require the caregiver to not only attend to the child, but also modulate feedback according to some complexity calculation, to the extent that these behaviors are not ensured (by innate or acquired filters and procedures in the caregiver), then further variation in outcomes should ensue.

Thus, for several reasons, all FB variants (ILN-FB, BuP-FB, and BuLP-FB) are most compatible with the observation of variation in outcomes, and thus are not plausible given current results.

**3.5 Summary.** This section classified phonological acquisition theories into five groups, which vary in their theoretical features (Table 1). Arguments were provided showing that, as a result of these features, some of these theories predict no variation in outcomes, others predict weak variation, and yet others are most compatible with considerable variation.

*Table 1. Classes of theories accounting for phonological acquisition, ranked in order of input quantities required (least to most), and key features of their Filters, Priors, and Outputs leading to this ranking. Prediction: Degree to which different classes of phonological acquisition theories predict outcome variation given input quantity variation, and main reason why. ILN stands for innate language networks, BuP for bottom-up phonology, BuLP for bottom-up lexicon and phonology, -SG for socially-gated variants, and -FB for feedback-based variants.*

| Class | Filters | Priors | Outputs | Prediction |
|---|---|---|---|---|
| ILN | All data | Representations & acoustic-representation links | Contrasts that are present | None (unless lexicon is used to determine phonological relevance) |
| BuP | All data | Primate sensitivities (& cross-modal info) | Distributions of sounds, sociophonetics | Some (sociophonetics is part of outputs, phonological categories extracted from environmental speech) |
| BuLP | All data | Primate sensitivities (& cross-modal info) | Distributions of sounds, sociophonetics, lexicon | More (lexical acquisition is part of outputs, stable phonological categories extracted from lexicon) |
| -SG | Only 1-on-1 data | Any of the above | Any of the above | Same as theories above (if there is the same degree of input variation for overall and directed input) |
| -FB | Only data with tutor feedback | Any of the above | Any of the above | Most (caregiver reinforcement, and thus interaction, is required) |

Given current evidence of large input differences and small outcome differences, it would appear that only the non-lexicon-based versions of ILN theories and potentially BuP theories (regardless of whether they are socially gated or not) are plausible. All other theories

(in the lexicon-based ILN group, in the BuLP group, and all FB variants) do not match well the observed pattern of considerable input variation and little or no outcome variation.

## 4. Generalizations: Features of theories that impact quantity requirements

One limitation of Section 3 is that the focus was on theories mainly concerned with early phonological development. However, a current trend is to contemplate language acquisition more holistically. At the very least, this entails considering how other linguistic levels emerge at the same time; and at most, such approaches assume that acquisition at a given level will impact some or all others (e.g., Frank, Goodman, & Tenenbaum, 2009; Johnson, Christophe, Dupoux, & Demuth, 2014).

It is therefore important to generalize the proposed meta-theoretical framework to data requirements as a function of theoretical features more broadly. To do so, I now re-visit each feature, and discuss how decisions regarding that feature affect data quantity requirements. As a result, such data quantity generalizations apply to all theories with that feature – regardless of where the theory sides on the emergentist versus nativist, domain-specific versus general, asocial versus social, phonology-only versus multi-level, or sequential versus parallel/interactive debates.

**4.1 Filters' specification.** Filters are what allows/forces information present in the environment to enter the processing workspace, and directs this information to specific procedures. Here, I will discuss two types of filters: Filters bearing on information processing in general, and filters that relate to social factors specifically. These two work extremely differently, and are thus addressed in two separate subsections.

**4.1.1 Informational filters.** Regarding general information processing, two filters have received some attention in the literature: a "speech bias" (e.g., Vouloumanos & Waxman, 2014) and an "infant-directed speech preference" (Frank et al., 2017). In the

context of the present framework, these are **filters** because they allow/force specific information present in the environment into the processing space: Infants attend preferentially to *that* aspect of their auditory experiences (i.e., attend more to speech than noise, attend more to a person talking in infant-directed speech than another talking in an adult-directed register). Authors interested in these filters often use them in the context of heuristics, further associating the processing of that signal with some specialized procedure; for example, Vouloumanos and Waxman (2014) argue that infants are more likely to extract an abstract pattern from speech than from non-speech, implying that pattern-finding procedures operate preferentially on speech. Another example: Csibra and Gergely (2009) claim that an infant-directed speech intonation leads infants to process a spoken word using interpretative rather than associative procedures.

In general terms, ILN theories as described in Section 3.1 rely on a linguistic filter – i.e., a filter that instantiates domain-specificity in terms of the input to the learning process. However, from the point of view of linguistics, there is no way to assume an innate filter that performs this chore, because linguistic material does not have stable surface features. There have been some proposals for physical definitions spanning modalities based on energy across frequencies (Poeppel, 2014). However, to prove this, it is necessary to program a computational algorithm that extracts information from human infants' input at this frequency from the auditory, visual, and tactile modality, and show that it ends up with all and only linguistic data.

An additional issue is that this type of broad definition will almost always be insufficient. To illustrate this, let us consider learning of phonetic categories from speech. The simplest, non-joint learning model proposed assumes that children attend to auditory input, and separate human speech from all background noise (speech preference). They then need to split this continuous speech signal into its constituent elements, which for the purposes of learning sound categories are phone-like units, and they then perform

second-order analyses to cluster these units into groups (Versteegh et al., 2015). This may be the only step in language acquisition that can operate on a single input representation. Even within phonological acquisition, there exist competing accounts, whereby phonological units are actually derived from an analysis that takes into account word identity (Swingley, 2005). Under such an account, the phonological acquisition system would perform an analysis of both the phonetics and the lexicon. More generally, for every linguistic level, there could exist a theory or model of acquisition that proposes that lexical knowledge plays a role (e.g., E & J., 2001 for grammar). Since lexical development will by necessity require multimodal input containing world representations external to language at the very least for all concrete words, then it follows that the notion of "input-domain specificity" is almost always inaccurate.

Such general information filters have not received sufficient attention from theorists working on language and cognitive development, perhaps because we do not realize the extent of the problem the infant faces when trying to determine what kinds of input are relevant for a given skill, in view of the fact that they are also perfecting other skills at the same time. To illustrate this, I will massively simplify the task facing human infants and describe it as minimally requiring the infant to learn both naive physics and language. As a result, infants will need to parse their multimodal experiences and decide which aspects of these experiences are relevant for learning naive physics, and which for learning language.

Take the everyday example of a child who, during mealtime, holds a sippy cup, then bounces it off a tray table and onto the floor, all the while hearing her caregiver talk. The tactile properties of the cup, the noise it makes when it falls, and the way it bounces are all incredibly useful to build a naive physics model. In contrast, these properties are transitory features of the content of the word "cup" (because cups can be made of a different material and not bounce at all). Tactile properties are thus central to naive physics, and peripheral for lexical acquisition. We cannot just assume that children will ignore tactile properties for learning phonological representations because deaf-blind individuals can perceive sign

language through their hands, and thus both information on physical properties and linguistic signals travel in this medium. In short, a simple filter "ignore tactile information when learning language or recognizing words" does *not* work.

When trying to explain learning of both naive physics and language, both by typically-perceiving infants and deaf-blind ones, a filter that may work is a "social versus asocial switch": Stream separately social and asocial multimodal experiences; if a multimodal experience is asocial, then process the information using the naive physics acquisition procedures; if it is social, process it with the language acquisition procedures. Notice that positing this filter would require a theory to assume that the infant is innately endowed with knowledge of what are the surface features of social experiences; or that they are born with procedures that allow them to learn such a filter (a review of social agent detection in Johnson, 2003).

Allowing all multimodal experiences to enter the acquisition workspace is equivalent to stating that deciding which aspects of the environment are informative is part of the acquisition process itself: Either the child is innately endowed with the "social versus asocial switch", or she has come to learn it over the course of acquisition. As a result, then, there is one more aspect of processing or knowledge that must be changed in the child. Put otherwise, learning the filter needs to be added to the learning **outputs**. As discussed elsewhere, the more the child needs to learn (i.e., the more is put in the **outputs**), the more data are required. Notice, incidentally, that such a theoretical move also calls for enriching the **procedures**, as now the theory must also specify how the filter is altered over the course of acquisition.

**4.1.2 Social filters.** I now turn to filters that relate to social interaction. Consider the following four filters; all else equal, a learner with Filter (1) will process more data in the environment than a learner with Filter (4):

1. The child learns from all environmental S (signal/speech/sign) (Rogoff, Paradise, Arauz, Correa-Chávez, & Angelillo, 2003)
2. The child learns only from child-directed, but not overheard, S (Shneidman, Arroyo, Levine, & Goldin-Meadow, 2013; Weisleder & Fernald, 2013)
3. The child learns only from adult child-directed S & own S when in one-on-one interaction (Ramírez-Esparza, García-Sierra, & Kuhl, 2014)
4. The child learns from adult child-directed S & own S when in one-on-one interaction, but only in child-initiated conversational blocks (extrapolating from Ramírez-Esparza et al., 2014)

It is an empirical question whether there is a quantity/quality tradeoff associated to this continuum: Perhaps processing overheard interaction means more data can get into the workspace, but these data are irrelevant to the child's current interests and thus do not capture the child's attention, or they are too difficult to process because they are spoken faster or less clearly, whereas interaction directly following the child's attention is obviously relevant to the child and (if the caregiver is cooperative) may be tailored to the child's processing abilities (e.g., Golinkoff, Can, Soderstrom, & Hirsh-Pasek, 2015). Regardless of whether it is ultimately beneficial for the child to do so or not, Filters (2-4) lead the learner to throw away all observed, third-party interactions.

Another open empirical question is which of these filters are active, and whether this changes as a function of what the child is learning, the child age, or even the child's environment. For instance, perhaps children exposed to highly cooperative tutors who always respond to the child's attention bids will come to settle on Filter (4). One way to test this involves setting up naturalistic third-party interactions that the infant can observe, but which are clearly not geared towards the child, and assessing whether the infant has learned from them. This issue has been studied mainly for word and action learning (see summary and discussion of whether Filter (4) may be innate as opposed to learned in Shneidman &

Woodward, 2016), although often this work compares the size of the effect in directed as opposed to overheard/third-party interaction. However, to test whether one of these filters is in place, one should establish the size of the learning effect in the third-party case. That is, if children only learn sounds, words, and/or morphosyntax from directed interaction, then they should fail systematically to uptake information on these levels present in third-party interaction.

**4.2 Priors' specification.** Since priors are innate knowledge, theories that rely on priors will, all else equal, be less dependent on data, since learners can fall back on that prior knowledge when data are sparse or inconclusive (provided that priors are helpful and match the infant's native language). This idea has often been discussed in association with nativist theories of syntax acquisition (see Pearl & Lidz, 2013 for sample discussion).

However, it is important to notice that, since priors can also be domain-general (see Table 2), the same intuition should apply to other types of theories relying on priors that are not domain-specific, including Bayesian theories in which priors are experience-independent probability distributions (extrapolating from Pearl, 2019), or to theories proposing that languages of the world tend to exploit features that are salient to our mammalian/primate brains (Christiansen & Chater, 2008). Thus, theories with strong priors will have a higher chance of leading to a functional language system even when provided with relatively little data (see Goldin-Meadow, 2014 for cases of zero input).

*Table 2. Examples of potential priors in language. Only domain-general examples are provided, but priors can also be posited that are language-specific (such as prior knowledge of possible human language structures).*

| Prior | Explanation |
|---|---|
| *Agents minimize effort* | This prior leads a phonological learner to expect that vowels may be centralized in high-frequency words. In social cognition it may be invoked to explain the inference that if an actor presses a button with his head when his hands are free, then perhaps this button only responds to head presses. |
| *Process wholes before parts* | This prior could follow from ease of processing imposed on by the visual system. In lexical acquisition, this prior would explain why children attach words to whole visual referents rather than subparts of the visual referent available (e.g., a whole animal rather than the animal body parts). |
| *Pay attention to ends* | This prior is akin to the recency effect found in non-verbal memory tasks, which is likely due to memory constraints. |

Among theories with priors, there is also an interesting prediction that "unnatural" patterns, meaning patterns that are rare in the world's languages or that are arbitrary, can appear and be maintained only in settings and cultures where children receive a great deal of input.

**4.3 Learning outputs' specification.** By and large, the more that has to be learned (by some endpoint), the more data will be needed. For instance, theories that posit as learning outputs the language's sound inventory will be less data-hungry than those supposing that the child must learn the language's sound inventory, syllabic structures, foot structures, and possible word structures.

Similarly, more input is required to learn conditional variants; that is, more input is required to learn separate sound inventories for different types of speakers (i.e., the sound

inventory is conditional on the speaker's individual and/or social identity). To illustrate this in a more specific way, consider the work summarized by Pierrehumbert (2016) as showing that both speakers and listeners expect different phonetic targets as a function of the talker's social identity, including gender, sexual orientation, social class, and social groups like "jocks" and "gangs" found in the same school. To keep track of such socially-contextual variants, learners need more data than learning a unique set of phonetic targets because the learner has to develop some form of classifier for each of those contexts (i.e., a classifier for gender, sexual orientation, social class, and jock/gang). In addition, for theories assuming that learners start out with differentiated streams and modalities for linguistic and social information (such as Dehaene-Lambertz & Gliga, 2004), more data will be needed to find such context-specific distributions than context-general ones. This is not true for theories that assume that the initial state involves holistic representations.

**4.4 Procedures' specification.** Procedures are what, taking into account priors (if they exist), process the information passed on by the filters in order to produce the learning outputs. Procedures could vary along many dimensions. Some of them will not be discussed here because they have already been covered in detail elsewhere, namely formal descriptions pertaining to the debate on poverty of the stimulus (e.g., Clark & Lappin, 2010; Magri, 2013); grammar complexity gradients (e.g., Heinz, 2016); and the value of positive versus negative, direct versus indirect evidence (e.g., Heinz, 2016; Pearl & Lidz, 2013). One procedure aspects that does not impact data requirements in a straightforward way and thus will not be discussed further: Whether the procedures themselves are domain-specific or domain-general (see Appendix A.3 for reasoning). In what remains, I discuss a few aspects that probably impact data requirements, although in all cases I will conclude that further formal, computational, and empirical work is necessary to establish this.

#### *4.4.1 Formal implementation.*

One may wonder to what extent formal implementation matters; that is, whether

learning occurs via constraint re-ranking (as in optimality theoretic or harmonic grammar approaches, Smolensky & Legendre, 2006), versus probability updates (Yang, 2002), versus learning weights of connections between units (in a neural network, Frank, Mathis, & Badecker, 2013). Upon first consideration, once priors and output complexity are accounted for (see Sections 4.2 and 4.3), then these details may not matter. However, a welcome future theoretical endeavor would be to assess whether, under reasonable assumptions, all of these are formally equivalent, in the footsteps of Goldwater, Griffiths, and Johnson (2009) showing that minimal description length systems are equivalent to unigram Bayesian descriptions for word segmentation problems.

### *4.4.2 Incremental versus batch learning.*

Another aspect of implementation is whether learners process their input incrementally versus in a batch mode. For the purposes of the present discussion, the question I am asking is exclusively one of order of information and decisions as follows: Incremental processing is that which attempts to extract information from the current experiences; any information that is not extracted on the fly is lost. In constrast, batch processing involves revisiting a given experience several times, allowing storage of information beyond the present experience.[2] Batch and incremental processing both predict better results with larger input quantity but for different reasons. For incremental processing, more data means better learning because the risk of missing criterial information is smaller over a larger body of experience. For batch processing, since experiences are processed multiple times, the risk of overfitting to the data is smaller for a larger body of experience. Moreover, direct comparisons between specific implementations of incremental compared to batch models of phonotactic acquisition (Magri, 2014) and word form discovery (Phillips & Pearl, 2015) have

[2]Animal learning results suggests that experiences are re-visited and re-parsed during sleep (De Lavilléon, Lacroix, Rondi-Reig, & Benchenane, 2015), in a process that ressembles batch re-processing of information. Thus, the conclusion that children cannot plausibly perform batch-style processing seems premature, and it is worth reasoning about whether one or the other style matter for data quantity.

surprisingly shown that the former can outperform the latter. However, this result bears on the final state performance rather than the learning curve - still leaving the question of incremental versus batch processing predictions for outcomes given input differences wide open.

#### *4.4.3 Parallel versus joint learning.*

Some theories propose that the acquisition of e.g. phonology, the lexicon, and morphosyntax feed into each other (joint models), whereas others assume that each level is acquired separately (either sequentially or in parallel), with minimal or no interaction. For instance, Kuhl (2007) describes infants' acquisition of sounds as impacting lexical learning, thus implicitly proposing a sequential order whereby the bases of phonology are acquired prior to, and at least initially independently from, the lexicon. In contrast, in Werker and Curtin (2005)'s account, infants learn both sounds and holistic words, with bidirectional effects between phonological and lexical representations; this is thus a joint interactive model.

Theories assuming joint learning of multiple linguistic levels face a similar issue as that discussed in the context of richness of the outputs: Joint learning implies a more complex model, which means that more interactions need to be tracked, and thus more data should be needed for equal precision. This leads to the conclusion that, all else equal, joint learning models will require more data. However, there are some counterexamples in the computational literature. In particular, Mark Johnson has investigated the process of building a lexicon of word forms from an unsegmented corpus (word segmentation), using various computational models with more or less joint-learning structure. Word segmentation results improve when the model both tries to build this lexicon *and* learn co-occurrences between words in English (e.g., Goldwater et al., 2009) but not in Sesotho (Johnson, 2008). One possibility is that joint learning improves outcomes provided that the noise across the levels is uncorrelated; perhaps learning co-occurrences is more helpful in English than Sesotho because the former has a relatively simpler morphophonology that renders

co-occurrence frames easier to find.

There is also another, unrelated way in which joint learning models may be more data-efficient than parallel or sequential models, and that is when procedures are set by priors cutting across linguistic domains. A clear example of this comes from generative grammar theories assuming innate parameters binding phonology and grammar: If the language is stress-initial (i.e., words tend to begin with a strong syllable followed by a weak one), then it will be morphosyntactically head-final (i.e., verb phrases will tend to have the complement followed by the verb rather than vice versa, Nespor et al., 2008). In this case, grammar learning requires no additional data, since the parameter at the morphosyntactic level is set when the child experiences phonological evidence (i.e., the learning outputs are attained in the complete absence of directly relevant data).

#### *4.4.4 Prediction versus compression.*

Finally, one dimension that should be considered when the present framework is extrapolated to lexical and morphosyntactic acquisition is whether there are differences in data sensitivity between algorithms based on prediction (e.g., Havron, Carvalho, Fiévet, & Christophe, 2019), as opposed to algorithms based on compression (e.g., Christiansen & Chater, 2016). Conceptually, the former are based on a generative or similar model, which constantly generates expectations regarding the next input unit, taking into account local context as well as patterns derived from previous experiences. Learning is then driven by prediction errors, which lead to alterations to the patterns in the long-term storage component. In contrast, compression-driven learning mechanisms try to represent the incoming signal in an economical manner. I suspect it will be particularly useful to revisit this in the context of lexical and morphosyntactic acquisition, where these two strategies may differ more in their predictions than they do in phonology.

**4.5 Summary.** In this section, a more detailed discussion of the proposed meta-theoretical framework allowed me to clarify which theoretical features determine quantity of data requirements in language acquisition more generally, including for levels other than the phonology and for theories considering multi-level acquisition. For these general cases, I have arrived at the conclusions that, all else equal, less data are needed...

- ... the more specific **information filters** are (see Section 4.1.1)
- ... the less specific **social filters** are (see Section 4.1.2)
- ... the more knowledge is incorporated as **priors** (see Section 4.2)
- ... the simpler the **outputs** (see Section 4.3)

Finally, Section 4.4 suggested that more work is needed to understand how certain features of procedures, such as whether linguistic levels are learned jointly or not, may affect data quantity requirements. That said, it is clear that any theory that relies crucially on lexical acquisition will be data-hungry (see Section 3.1.1), and thus predict individual, socioeconomic, and cultural variation as a function of input variation.

## 5 Discussion and future directions

This paper laid out a meta-theoretical framework which revealed that theories of early phonological acquisition vary widely in terms of the predictions they make regarding resilience to variation in input quantity. If we believe current data, children's input varies markedly, with estimates suggesting that children getting the most directed input hear 3-6 times as much as speech as children who are talked to the least. In contrast, differences in phonological development were not reliably detected. This would suggest that only a small class of theories is compatible with current data: Only theories relying on properly defined informational filters, strong priors, and well-specified learning procedures (assuming all of these are appropriate to the language being learned), where the learning procedure

contemplates only phonological acquisition and not also the lexicon. Given current results on laboratory learning suggesting that only child-directed interaction impacts phonological learning (Kuhl, Tsao, & Liu, 2003), it appears reasonable that the class of theories accommodating current results is even more circumscribed to the socially-gated variants of such theories. In contrast, theories where phonological acquisition depends at least partially on the lexicon, and those where caregiver contingent feedback is crucial, both predict individual and group differences as a function of input quantity (as well as quality, not addressed in this paper), and thus do not appear compatible with the current state of affairs. These interesting conclusions invite further empirical and theoretical research.

**5.1 Revising the empirical descriptions of the levels of input and outcome variation, and their association.** Given how surprising our conclusions are, perhaps the first line of research to recommend is to revisit the crucial empirical premises: That input variation is large, that outcome variation in phonological acquisition is small, and that outcome variation is uncorrelated to input variation.

Studies quantifying input and/or measuring phonological skills in infants are so costly that typically sample sizes collected by a single researcher are very small. For perception, the average sample size in non-native vowel discrimination experiments (Tsuji & Cristia, 2014) present on MetaLab (Bergmann et al., 2018) is about 17 infants. For production, most databases in PhonBank's North American English section contain data for fewer than 7 children (Rose & MacWhinney, 2014). Given these small sample sizes, most studies that report variation do so on the basis of qualitative inspection, rather than quantitative analyses. Combining across datasets in published experiments to assess the extent to which input variation is reflected in outcome variation will be challenging, since proxies of input variation (e.g., maternal education) are seldom reported. Fortunately, recent years have seen a rise in collaborative efforts, including collaborative data collection as in ManyBabies (Frank et al., 2017), and collaborative data analyses such as Bergelson et al. (2019).

Using larger samples to increase statistical power is imperative especially to establish the extent of variation in phonological outcomes because here we are trying to establish whether an effect is null – i.e., that it is small enough to be considered negligible. So how small should it be for us to be convinced that input variation does not lead to variation in phonological acquisition? One possible approach is to take a level of language where it seems obvious that input quantity variation should matter, namely the lexicon, and measure the size of that effect. Current best estimates of the effect of socioeconomic status on lexical processing is r=.26 when 2,252 children aged 12-42 months are included (Scaff & Cristia, 2019). Thus, it would be desirable to carry out a study in which, for instance, infants of varied socioeconomic statuses are recruited and their non-native sound discrimination is assessed, as well as the quantity of speech addressed to them. The prediction would then be that the correlation between the phonological experiment and speech quantity should be smaller than that between the lexicon and parental socioeconomic status. Assuming the effect found is r=0, we would need about 70 infants to show that the confidence interval does not include r=.26, and thus that the effect for phonology is smaller than that for the lexicon. If narrower confidence intervals are desired, e.g. excluding r=.1, then even more participants should be tested. In addition to showing that the correlation between input quantity (or a proxy for it, such as maternal socio-economic status) and infant phonological skills is smaller than that found for input and lexical skills, it would also be necessary to show that the measure of phonological skills used is sensitive to individual variation. That is, it would be trivial to obtain a null correlation with an insensitive task. A meta-analysis of past research on infant individual variation suggests that sound discrimination correlates with lexical development at about r=.35 (Cristia et al., 2014). Let us then imagine a study that measures (a) input quantity; (b) lexical development; and (c) phonological development, and finds that lexical development correlates with input quantity and phonological development to a similar extent, but phonological development and input quantity do not correlate. Crucially, the measures of lexical and phonological development should be gathered using

similar methods (for instance, via a laboratory experiment). This avoids an obvious confound: That lexical development correlates with input quantity to a greater extent than phonological development because the child's vocabulary size is estimated via a parental questionnaire, which may itself be affected by the caregiver's language skills, whereas the child's phonology is measured directly via a laboratory test. The correlation between lexical and phonological development is necessary to support the hypothesis that the phonological development measure is sensitive to individual variation, and should be expected if improved phonological categories facilitate word learning and recognition (Kuhl, 2007).

**5.2 Revising our theories and models.** If high-powered, carefully designed studies continue to show little to no variation in phonological outcomes by the end of the first year of life, and a non-significant (near zero) correlation with input quantity, then we would be led to conclude, based on the reasoning in Section 3, that only theories with a strong innate component and little reliance on the lexicon are plausible. In such a setting, there are two alternative explanations and two outstanding issues that need to be discussed further.

The first alternative explanation could posit that what infants learn in the first year of life is so trivial that innate filters and priors are unnecessary, and more generally that any learning procedure would suffice for learning. In fact, a recent information-theoretic account by Mollica and Piantadosi (2019) places the level of phonemes close to the bottom of the chart in terms of how many bits of information are required to encode it, with their best guess being 750 bits. Only syntax requires fewer bits (their best guess placing this level at 697) bits, with word frequency, phonemic wordforms (what we called proto-lexicon in Section 3.2), and lexical semantics requiring between two and three orders of magnitudes more bits. Note, however, these are estimations of how much information is required to encode this level once it has been acquired, and not how much is needed to develop that same knowledge. Additionally, regardless of where one sets the bar in terms of what needs to be acquired, it is still the case that the classes of theories described in Section 3 will meet that bar at different

points in time. In such a case, it would be necessary to not only inspect phonological knowledge at the end of the first year, but also at previous points in development, to establish whether differences in development as a function of input are visible earlier on. To give some examples, previous work suggests that infants show a behavioral preference for their native language (e.g., Moon, Cooper, & Fifer, 1993) and they may exhibit specialized neural processing for their dialect (e.g., Cristia, Minagawa-Kawai, et al., 2014) even before 4 months of age. Thus, we could ask whether infants exposed to widely different quantities of input vary in such behavioral and/or neural markers. If they do not, then we would still be in the general scenario of no variation in outcome despite variation in input, which would lead us back to the same subclasses of theories (strong filters and priors, no dependence on the lexicon).

The second alternative explanation states that we cannot observe the rate of learning because the markers we have been looking at (reduced discrimination of non-native sounds, increased discrimination of native sounds, increase in the prevalence of well-formed canonical syllables) are subject to strong maturational constraints such that they are typically triggered at a certain age (with exceptions related to the infant's auditory and biological environment; Werker & Hensch, 2015). If all children, regardless of how much input they have, come to accumulate sufficient evidence by this point in time, then we cannot use those markers as learning indices. In such a case, then indeed comparison of infants who hear a great deal versus very little input will not be informative for deciding among theories. It is beyond the present paper's scope to lay out how one could go about theory adjudication in such a setting.

As for the final issues that need to be brought up, first and foremost, it is crucial to remember that positing something to be innate is a hypothesis, and not a final scientific explanation. If we assume that infants have a specific innate informational filter, prior, or procedure, we must then explain how humans (and perhaps other animals) came to be

innately endowed with them, and how they are implemented and deployed in real life. Additionally, we need to demonstrate that a computational model implementing these precise innate components derives infant-like phonological knowledge when provided with realistic input (Dupoux, 2018).

Second, it would be ideal if theory proponents themselves developed more precise predictions made from their theories in terms of input quantity and quality that is necessary, sufficient, and/or optimal, further laying out what kinds of input-outcome relationship they expect (i.e., whethere it is a step function, linear, log-linear, and quadratic; see Section 5.4). This would be particularly welcome for usage-based theories (e.g., Ambridge & Lieven, 2015; McCauley & Christiansen, 2019), which give experience a determinant role in language development.

**5.3 Seeking stronger evidence for causation, and considering other language levels.** Correlational studies will be useful to represent natural variation in input and outcomes, and provided confounds are appropriately ruled out, a correlation between input and outcome that is essentially zero would be informative in the early phonological development scenario discussed above. However, both for later stages of phonological development (when a multiplicity of other factors start to play a role, including literacy, e.g., Horlyck, Reid, & Burnham, 2012) and for other levels of language in which we find correlations with input that are non-zero, correlational data is inherently problematic when attempting to assess causation. There has been a recent rise in interest for large-scale parental interventions centered in engaging the child in more conversation (see for instance the work by the LENA$^{TM}$ Foundation, https://www.lena.org/about/). Randomized control trials where the recommendation to engage their child with more talk is only made to some families (Weber, Fernald, & Diop, 2017) could be useful to establish a causal link between input quantity variation, on the one hand, and outcome variation, on the other. Specifically, if an intervention successfully modifies caregiver behavior by increasing the quantity of

speech the caregiver directs to the child (and changing nothing else in the caregiver's behavior), then comparison of children in the intervention versus control group in terms of, for example, lexical development and phonological development would provide more convincing evidence on a potential causal link between input and the lexicon, and lack thereof between input and phonology.

To state this most clearly, such a setting would be ideal to start delving into whether input affects different levels of language to different extents. Indeed, most (if not all) theories of early lexical development predict that child-directed, high quality input is a crucial engine spurring development (e.g., Smith & Yu, 2008; Trueswell et al., 2013; Yurovsky, 2018). This should lead to the prediction that the vocabulary should be highly sensitive to such interventions, with large differences between the intervention and control groups in this measure. If previous generalizations on phonological development hold true, one should find no difference between intervention and control groups on their phonological development. In contrast, other levels of language, notably morphosyntax, may provide intermediate levels of effects. For instance, work on emergent sign languages (Sandler, Meir, Padden, & Aronoff, 2005) and home signers (Goldin-Meadow, 2014; see also Goldin-Meadow, Brentari, Coppola, Horton, & Senghas, 2015) suggests that some properties of language are extremely resilient to the absence of input, whereas others are more dependent on it. As a result, the prediction would be that intervention and control group would differ for some but not all morphosyntactic measures.

It would be ideal in such a study to measure not only input quantity variation and a variety of outcomes tapping different linguistic levels, but also other aspects of caregiver behavior, to more clearly establish the causal pathways. That is, imagine that the intervention intends to get parents to increase the quantity of speech directed to the child, but in so doing also leads caregivers to be more sensitive to the child's focus of attention. As a result, a greater impact on, for instance, lexical than phonological development could be

explained by the fact that two aspects of caregiver have changed to improve conditions for lexical acquisition (input quantity and attention follow-in) but only one affects phonological acquisition (input quantity). Thus, even in a setting where causality can be better established than in correlational studies, it would be important to bear in mind potential confounds.

**5.4 Testing additional predictions.** In addition to discussing the main question of to what extent phonological outcomes are expected to vary as a function of input quantity variation, the present paper laid out a meta-theoretical model that allowed further predictions to be developed not only for phonology. Specifically, the expectation is that children who receive little input may learn some aspects of language but not others, and this at every level:

- In their phonology, they may not accumulate sufficient experience to acquire certain sound patterns (Chambers, Onishi, & Fisher, 2003).
- In their early lexicon, a higher proportion of words will be sound-symbolic or in some other way non-arbitrary (e.g., Imai et al., 2015).
- In their morphosyntax, there will be a stronger preference for unmarked word orders (Goldin-Meadow, 2014).

In addition to such qualitative differences, a finer-grained description of quantitative differences would be desirable. Throughout this paper, I have been discussing input variation in very broad terms, asking whether infants who get the most versus those who get the least input could vary in outcomes. This is sufficient if we find no differences in outcomes; but when we start looking at language levels where there is some outcome variation, then it would be informative to inspect the shape of this relationship.

To illustrate this, I drew data from three studies documenting input-outcome correlations as a function of socioeconomic status (Hart & Risley, 1995) or individual

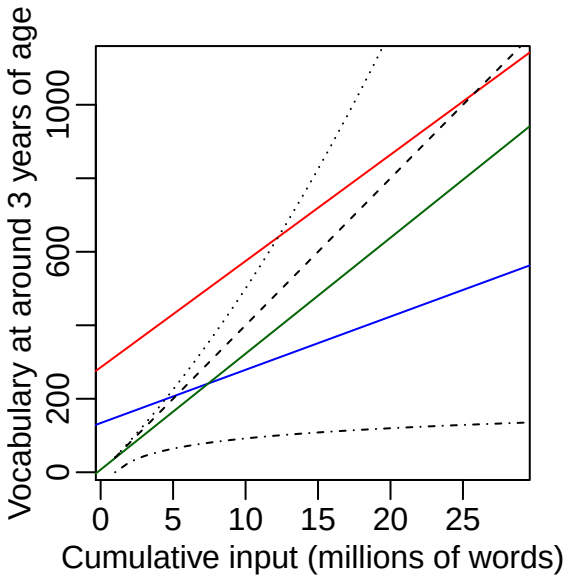


*Figure 1*. Vocabulary as a function of input. Solid lines represent linear regressions fitted to the input and vocabulary data reported in previous work (red: Hart and Risley, 1995; green: Shneidman et al., 2012; blue= Weisleder & Fernald, 2013). Each study had different methods to measure input and vocabulary, and thus differences in intercept should not be interpreted at all and differences in slope should be interpreted carefully, as they may also reflect methodological variation. Other lines represent simulated data with a simple linear fit (dashed), polynomial fit (dotted), and logarithmic fit (dot-dashed).

variation (Shneidman et al., 2013; Weisleder & Fernald, 2013). Each study used slightly different methods to measure input and outcomes (see Appendix A.4 for details). For input, Shneidman et al. (2013) counted words in one or two video-taped visits to the child, Hart and Risley (1995) drew averages from audiotapes gathered in a dozen visits, and Weisleder and Fernald (2013) used estimates extracted via an algorithm from a 16h nearly continuous recording. For outcomes, Shneidman et al. (2013) used the average of picture pointing and word elicitation tasks, Hart and Risley (1995) counted word types in the children's spontaneous production, and Weisleder and Fernald (2013) asked the caregiver to fill in a vocabulary checklist. Thus, we should not interpret differences in intercepts in this Figure. I fit a simple regression onto these data, a commonly used analysis in this line of work, with

all three studies reporting results of statistical analyses that assume a linear relationship between input and outcome. The three solid lines in Figure 1 demonstrate that the slope is clearly positive, although the Weisleder and Fernald (2013) line is shallower than the other two. This may mean that input explains less variance in outcomes in this group, although it could also be due to methodological differences across the studies.

Although most such work tests for linear relationships between input and outcome, one can imagine non-linear relationships that are cognitively plausible. I simulated data on a similar scale to illustrate this, by generating estimates of vocabulary using different formulas. Imagine that infants learn a word (in this vocabulary scale) for every 25,000 word tokens experienced – this linear relationship is conveyed by the dashed line. Alternatively, imagine that as children learn more words, they become better at learning words – this is a non-linear relationship with a positive second order component ($input + input^2$), represented here as the dotted line. Finally, imagine a different scenario, where additional exposure plays an increasingly smaller role, with the biggest gains being observed for initial exposure bouts – this is represented here by the dot-dashed line ($log(input)$). In all cases, a comparison of the lowest and the highest levels of exposure would reveal a difference across individuals or groups, and a simple regression assuming a linear relationship may not provide a worse fit for the data; but the intermediate levels of exposure can provide further evidence as to how precisely infants benefit from their experiences.

It would thus be important for future work to not only compare groups of infants varying in input quantity or fit statistics assuming a linear regression, but also consider the possibility that the relationship may be non-linear. Ideally, researchers will share input and outcome estimates at the individual level, so that we can eventually attempt meta- and mega-analyses (Costafreda, 2009) which may have more power to detect the precise shape of this relationship.

In sum, overall quantitative differences in language outcomes as a function of input

quantities, qualitative differences in what is learned, as well as a better understanding of the shape of the input-outcome relationship would all contribute to a better understanding of whether our theories of early language acquisition are on the right track in terms of what the role of the input is.

## 6 Conclusion

Variation in quantity of input is naturally present in cultures across the world, and even within the Western countries where most language acquisition data are collected. While the role of input had been discussed in the context of nativism versus emergentism and related debates, the current framework adds important insights by taking into account to a fuller extent key aspects of theory architecture, including the multimodal, multilevel structure of language acquisition. For the domain of phonological acquisition, this led to the realization that current results are best accommodated by a specific subclass of theories: Those that assume strong priors, strong informational filters, and do not rely on lexical acquisition. In addition, testable qualitative and quantitative predictions were drawn at multiple levels of language.

# Appendices

## A.1 Might children who hear less directed speech get more contextually-supportive speech?

Some previous discussions of cultural variation in input quantity appear to assume implicit compensatory mechanisms. For example, Lieven (2010) mentions that in cultures where talking to children is less common, caregivers produce more imperatives, which can provide good word-world correlation: "Don't stick your fingers in the socket" is a good learning environment for the real-world referent of "fingers" and "socket". Shneidman (2010) reports that Mayan caregivers, who speak less to their child, more frequently describe what the child is doing to a third party than American caregivers, who speak more to their child. Both of these examples are most compatible with a hidden assumption, that the caregiver has an internal "word-world correlation balance", and thus when the child is not provided with enough such data in directed, non-imperative interactions, the caregiver is drawn to producing such informative utterances in another way. This is akin to saying that the caregiver is producing a higher proportion of imperatives or descriptions of the child's activity *in order to* ensure language acquisition.

## A.2 A joint learner-teacher model

Some theories make reference to a tutor (e.g., the mother), who organizes the learner's language and world experience (e.g., Kuhl, 2007; Goldstein & Schwade, 2008; Gros-Louis et al., 2006; Yurovsky, 2018). Extrapolating from these, let us consider two extreme scenarios. At one extreme, the tutor just needs to use language to communicate with other people. Theories making this basic level of specification of the environment assume that children need only experience word-world correlations, and nothing else is needed for successful acquisition. At the opposite extreme, the tutor must be maximally cooperative: They talk

mainly (or only) about things and events that are the current focus of the child's attention (thus replacing the information filters); they hyperarticulate and produce sentences that are maximally informative and easy for the child to process (to ease the burden of the procedures). Notice that both extreme theories can accommodate a finding whereby children exposed to helpful tutors acquire language faster than children not benefiting from them: According to the basic theory, children proceed faster simply because the environment happens to be more informative. Thus, finding correlations between informative experiences and learning outputs cannot adjudicate between these theories. What would distinguish between such theories are cases in which the tutor is not cooperative if the children nonetheless learn. For instance, we should look for cases in which children are *never* addressed, and thus they must learn language from observed third-party interactions alone. If we were to find they nonetheless succeed, then theories positing that the acquisition process crucially depends on a helpful tutor would be shown to be incorrect. This is an old idea, but to my knowledge, no such cases can be verified independently and without any doubt (cf. Ochs & Schieffelin, 2001).

### A.3 Procedures' domain-specificity nature does not matter

Whether the operations applied onto the input to the acquisition system to result in language outcomes are shared with non-linguistic systems or not is a recurrent point of contention even among theories agreeing on the existence of linguistic priors, such as Principles-and-Parameters versus general Bayesian approaches (following Pearl, 2019's clear explanation). For instance, in Principles-and-Parameters approaches to syntactic acquisition, the learning procedures are assumed to be language-specific and not relevant to any other cognitive skill, because they make reference to linguistic units. In contrast, some Bayesian approaches express their priors using linguistic units, but assume that the procedure of acquisition itself is completely general, with the child updating their linguistic knowledge

using the same procedure that, say, birds use to update their knowledge of the likelihood of finding food in a given location. This example was selected to illustrate that whether a given procedure is shared with non-linguistic cognitive systems does not allow us to make any predictions as to whether the procedure will be more or less data hungry.

### A.4 Procedure to represent input-output relationships found in selected previous work

Input data for Hart and Risley (1995) comes from Figure 19, representing the estimated cumulative words addressed to the child. Although the caption does not detail how these numbers were derived, it does state they are based on the average number of word tokens found in the hours that were observed. Since the numbers are cumulated over the child's life, it is likely that they authors multiplied by awake time to get an estimate per day, and multiplied by the number of days needed to arrive at cumulative numbers up to age three years (yielding averages of 8, 16.5, and 28.5 million words for the three socioeconomic groups in that study). The outcome data also corresponds to age 3, and it was extracted from Figure 2, which represents the "average number of vocabulary words recorded in that and all prior months" (averaging 525, 749, and 1,116 cumulative word types for the three groups in that study).

The data for Weisleder and Fernald (2013) was drawn from Figure 3a using an online tool (https://automeris.io/WebPlotDigitizer/), yielding number of words in a vocabulary checklist and an average number of word tokens per hour addressed to the child. To adapt the input data to the same scale as Hart and Risley (1995), these averages were multiplied by 10 awake hours per day, 365 days per year, and three years. The outcome measure was not scaled.

Finally, Shneidman et al. (2013)'s Study 4 reported on the number of word types

directed to the child during the observations, which is a conceptually different input measure than that found in the other two studies. Therefore, we used instead the average and standard deviation in number of word tokens for the same 15 children's data, which were reported on p. 41 of Shneidman (2010). A random sample of 15 data points was generated using this mean and standard deviation. As with Weisleder and Fernald (2013), these estimates representing numbers of word tokens per hour were multiplied by 10 x 365 x 3 to arrive at an input estimated in millions of word tokens cumulated up to age three years. The outcome measures were derived using the regression estimate provided in Shneidman et al. (2013)'s Study 4 (unstandardized using the word tokens mean and standard deviation), multiplied by the standard deviation in Weisleder and Fernald (2013) and adding the mean from that same study so that it would be on the same scale.

**Acknowledgments**

I am grateful for the support of the Agence Nationale de la Recherche (ANR-17-CE28-0007 LangAge, ANR-16-DATA-0004 ACLEW, ANR-14-CE30-0003 MechELex, ANR-17-EURE-0017); and the J. S. McDonnell Foundation Understanding Human Cognition Scholar Award. I thank Emmanuel Dupoux, Teodora Gliga, Naomi Havron, Paul Iverson, Francis Mollica, Lisa Pearl, Steven Piantadosi, Amanda Seidl, Sho Tsuji, and anonymous reviewers for helpful discussion and/or feedback on previous versions of this work. I take full responsibility for the contents. Declarations of interest: none.

## References

Ambridge, B., & Lieven, E. (2015). A constructivist account of child language acquisition. In B. MacWhinney & O. W (Eds.), *The handbook of language emergence* (pp. 478–510). Hoboken, NJ: Wiley Online Library.

Bergelson, E., Casillas, M., Soderstrom, M., Seidl, A., Warlaumont, A. S., & Amatuni, A. (2019). What do North American babies hear? A large-scale cross-corpus analysis. *Developmental Science*, *22*(1), e12724.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, *109*(9), 3253–3258.

Bergelson, E., & Swingley, D. (2018). Young infants' word comprehension given an unfamiliar talker or altered pronunciations. *Child Development*, *89*(5), 1567–1576.

Bergmann, C., Tsuji, S., Piccinini, P. E., Lewis, M. L., Braginsky, M., Frank, M. C., & Cristia, A. (2018). Promoting replicability in developmental research through

meta-analyses: Insights from language acquisition research. *Child Development*, *89*(6), 1996–2009.

Chambers, K., Onishi, K., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition*, *87*, B69–B77.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, *31*(5), 489–509.

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*.

Clark, A., & Lappin, S. (2010). *Linguistic nativism and the poverty of the stimulus*. John Wiley & Sons.

Costafreda, S. G. (2009). Pooling fMRI data: Meta-analysis, mega-analysis and multi-center studies. *Frontiers in Neuroinformatics*, *3*, 33.

Cristia, A. (2019). Language input and output in early childhood: A systematic review. *https://acristia.shinyapps.io/speechInputXcultLongiCum*.

Cristia, A., Minagawa-Kawai, Y., Egorova, N., Gervain, J., Filippin, L., Cabrol, D., & Dupoux, E. (2014). Neural correlates of infant accent discrimination: An fNIRS study. *Developmental Science*, *17*(4), 628–635.

Cristia, A., Seidl, A., Junge, C., Soderstrom, M., & Hagoort, P. (2014). Predicting individual variation in language from infant speech perception measures. *Child Development*, *85*(4), 1330–1345.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148–153.

Curtin, S., Byers-Heinlein, K., & Werker, J. F. (2011). Bilingual beginnings as a lens for theory development: PRIMIR in focus. *Journal of Phonetics*, *39*(4), 492–504.

Dehaene-Lambertz, G., & Gliga, T. (2004). Common neural basis for phoneme processing in infants and adults. *Journal of Cognitive Neuroscience*, *16*(8), 1375–1387.

De Lavilléon, G., Lacroix, M. M., Rondi-Reig, L., & Benchenane, K. (2015). Explicit memory creation during sleep demonstrates a causal role of place cells in navigation. *Nature Neuroscience*, *18*(4), 493.

Dupoux, E. (2018). Cognitive science in the era of artificial intelligence: A roadmap for reverse-engineering the infant language-learner. *Cognition*, *173*, 43–59.

E, B., & J., G. (2001). On the inseparability of grammar and the lexicon: Evidence from acquisition. In B. E & G. J. (Eds.), *Language development: The essential readings* (pp. 134–162). Oxford, UK: Blackwell Publishers.

Eilers, R. E., Oller, D. K., Levine, S., Basinger, D., Lynch, M. P., & Urbano, R. (1993). The role of prematurity and socioeconomic status in the onset of canonical babbling in infants. *Infant Behavior and Development*, *16*(3).

Fourtassi, A., & Dupoux, E. (2014). A rudimentary lexicon and semantics help bootstrap phoneme acquisition. In *Proceedings of the eighteenth conference on computational natural language learning* (pp. 191–200).

Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., ... Yurovsky, D. (2017). A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. *Infancy*, *22*(4), 421–435.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science*,

*20*(5), 578–585.

Frank, R., Mathis, D., & Badecker, W. (2013). The acquisition of anaphora by simple recurrent networks. *Language Acquisition*, *20*(3), 181–227.

Frank, S., Feldman, N. H., & Goldwater, S. (2014). Weak semantic context helps phonetic learning in a model of infant language acquisition. In *Proceedings of the 52nd annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 1073–1083).

Goldin-Meadow, S. (2014). In search of resilient and fragile properties of language. *Journal of Child Language*, *41*(S1), 64–77.

Goldin-Meadow, S., Brentari, D., Coppola, M., Horton, L., & Senghas, A. (2015). Watching language grow in the manual modality: Nominals, predicates, and handshapes. *Cognition*, *136*, 381–395.

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, *19*(5), 515–523.

Goldwater, S., Griffiths, T. L., & Johnson, M. (2009). A Bayesian framework for word segmentation: Exploring the effects of context. *Cognition*, *112*(1), 21–54.

Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby) talk to me: The social context of infant-directed speech and its effects on early language acquisition. *Current Directions in Psychological Science*, *24*(5), 339–344.

Gonzalez-Gomez, N., & Nazzi, T. (2012). Phonotactic acquisition in healthy preterm infants. *Developmental Science*, *15*(6), 885–894.

Gros-Louis, J., West, M. J., Goldstein, M. H., & King, A. P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of*

*Behavioral Development*, *30*(6), 509–516.

Hale, M., & Reiss, C. (2003). The subset principle in phonology: Why the tabula can't be rasa. *Journal of Linguistics*, *39*(2), 219–244.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children.* Baltimore, MD: Paul H Brookes Publishing.

Havron, N., Carvalho, A. de, Fiévet, A.-C., & Christophe, A. (2019). Three-to four-year-old children rapidly adapt their predictions and use them to learn novel word meanings. *Child Development*, *90*(1), 82–90.

Hayes, B. (2004). Phonological acquisition in optimality theory: The early stages. In R. Kager, J. Pater, & W. Zonneveld (Eds.), *Constraints in phonological acquisition* (pp. 158–203). Cambridge, UK: Cambridge University Press.

Heinz, J. (2016). Computational theories of learning and developmental psycholinguistics. In J. Lidz, W. Synder, & J. Pater (Eds.), *The Oxford Handbook of Developmental Linguistics* (pp. 633–663). Oxford, UK: Oxford University Press.

Horlyck, S., Reid, A., & Burnham, D. (2012). The relationship between learning to read and language-specific speech perception: Maturation versus experience. *Scientific Studies of Reading*, *16*(3), 218–239.

Imai, M., Miyazaki, M., Yeung, H. H., Hidaka, S., Kantartzis, K., Okada, H., & Kita, S. (2015). Sound symbolism facilitates word learning in 14-month-olds. *PLoS One*, *10*(2), e0116494.

Johnson, M. (2008). Unsupervised word segmentation for Sesotho using adaptor grammars. In *Proceedings of the Tenth Meeting of the Association for Computational Linguistics Special Interest Group on Computational Morphology and Phonology* (pp. 20–27).

Johnson, M., Christophe, A., Dupoux, E., & Demuth, K. (2014). Modelling function words improves unsupervised word segmentation. In *ACL* (pp. 282–292).

Johnson, S. C. (2003). Detecting agents. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1431), 549–559.

Jusczyk, P. W. (1993). From general to language-specific capacities: The WRAPSA model of how speech perception develops. *Journal of Phonetics*, *21*, 3–28.

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, *33*(5), 630–645.

Kaplan, F., Oudeyer, P.-Y., & Bergen, B. (2008). Computational models in the debate over language learnability. *Infant and Child Development: An International Journal of Research and Practice*, *17*(1), 55–80.

Kuhl, P. K. (1991). Human adults and human infants show a perceptual magnet effect for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*, *50*(2), 93–107.

Kuhl, P. K. (2007). Is speech learning 'gated' by the social brain? *Developmental Science*, *10*(1), 110–120.

Kuhl, P. K., Tsao, F.-M., & Liu, H.-M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, *100*(15), 9096–9101.

Lenneberg, E. H. (1969). On explaining language. *Science*, *164*(3880), 635–643.

Lieven, E. (2010). Language development in a cross-linguistic context. In M. Kail & M. Hickmann (Eds.), *Language acquisition across linguistic and cognitive systems* (pp. 91–108). Amsterdam: John Benjamins.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk. 3rd Edition.* Mahwah, NJ: Lawrence Erlbaum Associates.

Magri, G. (2013). The complexity of learning in optimality theory and its implications for the acquisition of phonotactics. *Linguistic Inquiry*, *44*(3), 434–468.

Magri, G. (2014). Error-driven versus batch models of the acquisition of phonotactics: David defeats goliath. *Proceedings of the Annual Meetings on Phonology*, *1*(1).

Majorano, M., Vihman, M. M., & DePaolis, R. A. (2014). The relationship between infants' production experience and their processing of speech. *Language Learning and Development*, *10*(2), 179–204.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can effect phonetic discrimination. *Cognition*, *82*, B101–B111.

McCauley, S. M., & Christiansen, M. H. (2019). Language learning as language use: A cross-linguistic model of child language development. *Psychological Review*, *126*(1), 1.

McGillion, M., Herbert, J. S., Pine, J., Vihman, M. M., DePaolis, R., Keren-Portnoy, T., & Matthews, D. (2017). What paves the way to conventional language? The predictive value of babble, pointing, and socioeconomic status. *Child Development*, *88*(1), 156–166.

Melvin, S. A., Brito, N. H., Mack, L. J., Engelhardt, L. E., Fifer, W. P., Elliott, A. J., & Noble, K. G. (2017). Home environment, but not socioeconomic status, is linked to differences in early phonetic perception ability. *Infancy*, *22*(1), 42–55.

Mollica, F., & Piantadosi, S. T. (2019). Humans store about 1.5 megabytes of information during language acquisition. *Royal Society Open Science*, *6*(3), 181393.

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language.

*Infant Behavior and Development*, *16*(4), 495–500.

Nespor, M., Shukla, M., Vijver, R. van de, Avesani, C., Schraudolf, H., & Donati, C. (2008). Different phrasal prominence realizations in VO and OV languages. *Lingue E Linguaggio*, *7*(2), 139–168.

Ochs, E., & Schieffelin, B. (2001). Language acquisition and socialization: Three developmental stories and their implications. In A. Duranti (Ed.), *Linguistic anthropology: A reader* (pp. 263–301). Oxford, England: Blackwell.

Oller, D. K., Eilers, R. E., Basinger, D., Steffens, M. L., & Urbano, R. (1995). Extreme poverty and the development of precursors to the speech capacity. *First Language*, *15*(44), 167–187.

Oller, D. K., Eilers, R. E., Steffens, M. L., Lynch, M. P., & Urbano, R. (1994). Speech-like vocalizations in infancy: An evaluation of potential risk factors. *Journal of Child Language*, *21*(1), 33–58.

Pace, A., Luo, R., Hirsh-Pasek, K., & Golinkoff, R. M. (2017). Identifying pathways between socioeconomic status and language development. *Annual Review of Linguistics*, (3), 285–308.

Pearl, L. S. (2019). Modeling syntactic acquisition. (J. Sprouse, Ed.). Oxford, UK: Oxford University Press.

Pearl, L. S., & Lidz, J. (2013). Parameters in language acquisition. In C. Boeckx & K. Grohmann (Eds.), *The Cambridge Handbook of Biolinguistics* (pp. 129–159). Cambridge: Cambridge University Press.

Petitto, L. A., Holowka, S., Sergio, L. E., Levy, B., & Ostry, D. J. (2004). Baby hands that move to the rhythm of language: Hearing babies acquiring sign languages babble

silently on the hands. *Cognition*, *93*(1), 43–73.

Phillips, L., & Pearl, L. S. (2015). The utility of cognitive plausibility in language acquisition modeling: Evidence from word segmentation. *Cognitive Science*, *39*(8), 1824–1854.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, *46*(2-3), 115–154.

Pierrehumbert, J. B. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics*, *2*(1), 33–52.

Poeppel, D. (2014). The neuroanatomic and neurophysiological infrastructure for speech and language. *Current Opinion in Neurobiology*, *28*, 142–149.

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2014). Look who's talking: Speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science*, *17*(6), 880–891.

Räsänen, O. (2012). Computational modeling of phonetic and lexical learning in early language acquisition: Existing models and future directions. *Speech Communication*, *54*(9), 975–997.

Rogoff, B., Paradise, R., Arauz, R. M., Correa-Chávez, M., & Angelillo, C. (2003). Firsthand learning through intent participation. *Annual Review of Psychology*, *54*(1), 175–203.

Rose, Y., & MacWhinney, B. (2014). The PhonBank Project: Data and software-assisted methods for the study of phonology and phonological development. In J. Durand, U. Gut, & G. Kristoffersen (Eds.), *The Oxford Handbook of Corpus Phonology, 380?401* (pp. 633–663). Oxford, UK: Oxford University Press.

Sandler, W., Meir, I., Padden, C., & Aronoff, M. (2005). The emergence of grammar: Systematic structure in a new language. *Proceedings of the National Academy of*

*Sciences*, *102*(7), 2661–2665.

Scaff, C., & Cristia, A. (2019). Socioeconomic status and early lexical processing: A meta-analysis.

Shneidman, L. A. (2010). *Language input and acquisition in a Mayan village* (PhD thesis). The University of Chicago.

Shneidman, L. A., Arroyo, M. E., Levine, S. C., & Goldin-Meadow, S. (2013). What counts as effective input for word learning? *Journal of Child Language*, *40*(3), 672–686.

Shneidman, L., & Woodward, A. L. (2016). Are child-directed interactions the cradle of social learning? *Psychological Bulletin*, *142*(1), 1.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568.

Smolensky, P., & Legendre, G. (2006). *The harmonic mind: From neural computation to optimality-theoretic grammar*. Cambridge, MS: MIT press.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, *50*(1), 86–132.

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology*, *66*(1), 126–156.

Tsuji, S., & Cristia, A. (2014). Perceptual attunement in vowels: A meta-analysis. *Developmental Psychobiology*, *56*(2), 179–191.

Tsuji, S., Dupoux, E., & Cristia, A. (2019). A framework for social theories of language acquisition.

Versteegh, M., Thiolliere, R., Schatz, T., Cao, X. N., Anguera, X., Jansen, A., & Dupoux, E. (2015). The zero resource speech challenge 2015. In *Sixteenth annual conference of the international speech communication association*.

Vouloumanos, A., & Waxman, S. R. (2014). Listen up! Speech is for thinking during infancy. *Trends in Cognitive Sciences*, *18*(12), 642–646.

Warlaumont, A. S., Richards, J. A., Gilkerson, J., & Oller, D. K. (2014). A social feedback loop for speech development and its reduction in autism. *Psychological Science*, *25*(7), 1314–1324.

Weber, A., Fernald, A., & Diop, Y. (2017). When cultural norms discourage talking to babies: Effectiveness of a parenting program in rural Senegal. *Child Development*, *88*(5), 1513–1526.

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11), 2143–2152.

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, *1*(2), 197–234.

Werker, J. F., & Hensch, T. K. (2015). Critical periods in speech perception: New directions. *Annual Review of Psychology*, *66*, 173–196.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63.

Yang, C. (2002). *Knowledge and learning in natural language*. Oxford, UK: Oxford University Press.

Yurovsky, D. (2018). A communicative approach to early word learning. *New Ideas in Psychology*, *50*, 73–79.