Chapter 18 Beyond the CDI

Throughout this book, we have attempted to gain the broadest possible understanding of children's language learning by engaging deeply with data from the CDI family of instruments. We hope that readers agree that this exercise has been very fruitful in uncovering a variety of patterns that can inform our understanding of language learning. But, it has also uncovered a wide variety of limitations to the CDI, which in turn restrict the issues on which we can comment. In this chapter, we begin by discussing some methodological morals from this work for psychology more broadly. We then turn again to the limitations of the CDI to address the question of how language acquisition research can move beyond the CDI.

18.1 Methodological morals

As we noted in Chapter 1, psychology has recently been plagued by concerns about reproducibility (e.g., Hardwicke et al. 2018) and replicability (e.g., Open Science Collaboration 2015). Our work here was in part inspired by considering these issues and their impact on the field of language development. The ultimate goal of research in the area of language learning is to create a quantitative theory that allows for precise predictions and principled explanations of developmental phenomena (Dupoux 2018). Such a theory cannot be built on a series of non-reproducible findings and binary conclusions (Frank et al. 2017).

Wordbank is one reply to this situation: By compiling the extant CDI datasets into a single open database, researchers can reproduce previous and new research conclusions that use these data. The analyses we report here are computationally reproducible through the availability of the code necessary to build the book and all its figures and analyses. In addition, by seeking a level of scale beyond previous efforts, we have attempted to avoid the variability inherent in "small-N" studies.

Further, our work is built on the notion of replication. Nearly every one of the preceding chapters is in some sense a "replication" of previous work — an analysis from previous research with one CDI dataset was applied (sometimes with modifications) to other datasets (and languages). Yet, the result is not a judgement or referendum on the original; we do not declare a binary success or failure of the replication attempt — but if we did, our success rate would be very high! Instead, we are interested in the *degree* to which a particular quantitative estimate varies across languages and cultures.

This sort of analysis is superficially similar to the idea of "hidden moderators" that has plagued the replication debate (Van Bavel et al. 2016; cf. Inbar 2016). That line of explanation has been an attempt to contextualize failures to replicate particular experimental effects by invoking unknown sources of variability across contexts. In contrast, our efforts here allow us to quantify variation across "replications" of the same effect and use these estimates as the signal — rather than as noise to be discarded or averaged out.

One notable feature of our analytic strategy is that we rely very little on binary decision-theoretic inferences using null hypothesis significance testing. There are a handful of p-values throughout the book, but few of these license any prominent inferential conclusion; mostly they exist to provide a quick check that a particular slope is likely to be nonzero. Instead, our goal has been to measure quantities of interest with high precision, looking for statistical measures that relate to our theoretical goals. For example, the existence of a noun bias is a fascinating observation, but the observation alone gives limited leverage to differentiate theories. In contrast, the precise magnitude of a noun bias for a particular sample provides more leverage for quantitative theorizing. And the distribution of magnitudes across many of the world's languages gives greater leverage still.

Yet our analytic strategy is only as useful as the data it relies on, and these have substantial limitations. Some of these limitations are imposed by the specifics of our data, and others come from fundamental limitations of the CDI.

18.2 Limitations of Wordbank and the CDI

Reprising our discussion in Chapter 1, despite the large number of children represented in the Wordbank dataset, there are still a number of major omissions. First, because the data are typically from normative studies, data on atypical development are not represented, even though the CDI has been used profitably with several developmental disorder populations (Heilmann et al. 2005; Luyster, Lopez, and Lord 2007). Second, for similar reasons, bilingual data are almost entirely absent, though exciting work with multi-lingual children is beginning to emerge (Bilson et al. 2015; Floccia et al. 2018). Third, the number of children with consistent longitudinal observations is still relatively small. Although we used longitudinal data in Chapters 4, 13, 14, and 15, the extant data provide at best a limited picture of change over time within individuals and across languages, since our conclusions were drawn almost exclusively from English and Norwegian data.

Beyond these data availability issues, the CDI as an intrument is simply not the appropriate tool for asking every kind of question about child language development. Following the metaphor we introduced in Chapter 1, the CDI is a "macro-economic" indicator. It tells us about the global profile of a child's linguistic abilities, rather than revealing the local "micro-economic" dynamics of learning at a particular point in developmental time. The local dynamics of children's learning, language use in communication, and comprehension in the moment have all been important targets for empirical investigation (e.g., Clark 1988; Fernald et al. 1998; Smith and Yu 2008). At best, a CDI can provide some emergent average of these processes over time, much the same way the gross domestic product of a nation describes a summary of the impact of all the contributing markets.

The use of parent report to provide a global picture of the child's entire language system — from gestures to vocabulary and grammar — is also a weakness when it comes to addressing detailed questions about the representation of specific words. Because parents are not linguists, they cannot be profitably asked the kinds of targeted questions that might shed light on a variety of theoretical issues. For example, a tremendous amount of research has investigated the development of children's phonological systems (e.g., Vihman 1996). Research with the CDI must remain silent on this topic — we instruct parents to check "says" if the child produces *any* appropriate phonological form.

Similarly, an important target for research on vocabulary learning is the type of semantic generalizations that children make, including whether words are initially over- or under-generalized (e.g., Clark 1973) and how their appropriate extension is found (e.g., Markman 1990; Xu and Tenenbaum 2007). The CDI depends on parents to, on average, detect appropriate production and comprehension of words in specific contexts. But these averages of individual uses necessarily reveal relatively little about the nature of the semantic representations underlying the uses of the word — even for concrete objects, but especially for descriptive or closed-class words. A clear example of this phenomenon appears in Chapter 11. We see that time words are under-represented in children's vocabulary; but they are still present. Yet, according to Tillman and Barner (Tillman and Barner 2015; Tillman et al. 2017), 2.5-year-olds probably have incorrect or incomplete semantics for essentially all of these words. They still utter the words in appropriate contexts. If the semantics were probed more carefully, however, gaps with adult-like representations would become readily apparent. Experimental methods are likely to be more effective than parent report in these sorts of cases.

In sum, Wordbank and the CDI itself are the right tools for certain kinds of questions. But there are many other questions — some of which arise naturally in and from our work here — that cannot be addressed with these tools. How do we move forward beyond the CDI?

18.3   What comes next?

While parent report is no substitute for laboratory observations and experiments, new technical developments suggest that there may be ways to get traction on the questions discussed above by developing successors to the CDI. These approaches are typically inspired by the CDI and the potential of parent report, but they are also not limited by the design features of the specific assessment. We briefly discuss three promising directions for this type of work: adaptive testing, web–based assessment, and app–based assessment.

18.3.1   Web–based assessment

CDI forms have traditionally been administered to parents on paper; in general, the CDI community has been relatively slow to transition to online administration despite the prevalence of online methods in experimental and survey research during the last ten years (Buhrmester, Kwang, and Gosling 2011). One major exception has been two Northern European normative datasets included in Wordbank, namely the Norwegian and Danish datasets. As discussed in Kristoffersen et al. (2013), with the creation of an appropriate administration system, these groups were highly successful in recruiting large samples of parents to complete the CDI, and their experiences suggest that it is possible to get high quality data — even deep longitudinal data — through online administration.

Adoption of web–based methods throughout the CDI community has thus far been spotty, however. One obstacle is the varying copyright status of different CDI instruments, which creates legal roadblocks to a simple adaptation of the instrument to a uniform online format. Our personal experience is that some research groups ignore these restrictions and create their own ad–hoc surveys using platforms like Google Surveys, Survey Monkey, or Qualtrics. These ad–hoc versions often do not have consistent administration instructions or formatting, and sometimes contain errors or omissions as well, due to the challenges of porting form content to a new format.

To address these challenges, we have created `web-cdi` (http://web–cdi.stanford.edu), an online platform that allows researchers to administer a growing range of CDI forms to participants by generating shareable hyperlinks. The system contains a study–management interface so that researchers can create batches of administrations that are specific to a particular participant group (and can reuse demographic data for multiple administrations to the same participant). At the time of writing, we are working to navigate the legal restrictions on a number of instruments so that web–cdi can be used by researchers studying a wide variety of languages. (Although we have not yet implemented adaptive testing in the web–cdi system (see also below), in principle, this method would be relatively straightforward to include as an "instrument" in the broader system.)

A flexible and general web–based administration system like web–cdi has the potential to facilitate efforts to broaden the Wordbank dataset. First, more representative segments of particular national populations can be reached by advertising through email and social media. Such efforts could substantially improve the generalizability of the demographic analyses reported in Chapter 6. Our group is currently experimenting with using social media advertising to recruit lower–income and racial/ethnic minority groups that are typically underrepresented in research on language development in the United States. Further, as shown by Kristoffersen et al. (2013), online administration substantially simplifies the creation of longitudinal datasets. In the years to come, we hope that other groups (including our own) can leverage web–cdi to create longitudinal datasets that span the full time period of the emergence of language. If they came from a typologically diverse sample of languages, such datasets would be especially valuable in generalizing our analyses of grammatical and morphological development in Chapters 13 and 14.

Finally, though there has been some important and impactful bilingual CDI research (e.g., Bilson et al. 2015; Floccia et al. 2018), substantial challenges remain in this domain. For example, in many areas, the population of bilinguals is very diverse and so the provision of the appropriate CDI forms with language–specific instructions is a non–trivial challenge. Further, the scoring of CDI forms across languages requires conceptual mappings across words (such as those used in Chapter 10), so that the child's total conceptual vocabulary — the number of concepts the child has names for across all languages, regardless of the language used — can be established (Core et al. 2013). Both of these challenges can be ameliorated by the design of an appropriate platform. Although the technical challenges are not insignificant, in principle, web–cdi can make it easy for researchers to share a wide variety of forms with parents (with customized, language–specific instructions), gathering data remotely from a more diverse sample of bilinguals. Further, the platform can leverage cross–instrument mappings from our work here to provide measures of both total vocabulary and total conceptual vocabulary.

18.3.2 Adaptive approaches

Although we have made much use in our analyses of the extensive item list for the CDI, the long forms of the CDIs are simply overkill for some applications. For researchers and clinicians who wish to recover a single percentile score for a child's overall vocabulary size — for example, for purposes of language screening — the CDI is far too long. To ask parents to make hundreds of responses typically takes from ten minutes to upwards of a half an hour, and the reliability coefficients across words are high enough that the forms can be shortened substantially without losing assessment fidelity for determining a child's overall vocabulary size.

To address this issue, researchers have developed CDI short forms (e.g., L. Fenson, Pethick, et al. 2000; Jackson–Maldonado, Marchman, and Fernald 2013). These forms typically contain around 100 words, chosen for their ability to distinguish children across a range of abilities at both younger and older ages. Like the long forms, there is substantial evidence for their validity (e.g., Can et al. 2013). Indeed, some studies even have used so–called "short–short" forms with as few as 25–50 words, chosen to get a rough measure of variation in vocabulary size using much shorter vocabulary lists (Andreassen and Fletcher 2007). We have not focused on including data from short forms in Wordbank, in part because they provide relatively little traction on issues of vocabulary composition and relations between parts of the language system (issues that are critical to our syntheses in Chapters 16 and 17). But, for many users, they are more efficient and appropriate instruments than the long–form CDIs for obtaining an overall score that reflects a child's standing relative to same–aged peers.

One exciting possibility is to achieve further gains in efficiency in the context of web– or app–based administration through the use of adaptive methods. In these sorts of methods, which are common in educational assessment settings, words are chosen to provide maximal information about the child's place in the ability distribution. For example, Makransky et al. (2016) used item–response theory simulations to demonstrate that an adaptively–chosen set of 50 words could in principle recover percentile ranks with a correlation of .95 with the full sample. A correlation of .85 could be achieved with as few as 10 words, provided that they were chosen appropriately. And using a computational approach based on the use of Wordbank data for the creation of empirical norms, Mayor and Mani (2018) showed good performance in simulation and an empirical validation using the German FRAKIS CDI (Szagun et al. 2006).[34] Thus, adaptive methods have substantial promise for quick global assessments. Further, adaptive methods that aim for global vocabulary estimates could be paired with in–depth questions about specific semantic domains or word classes for an efficient approach to exploring particular theoretical issues.

18.3.3 App—based approaches

A final promising direction for parent reports about children's language is the use of mobile apps. Mobile technology has already been used productively for eliciting emotional information through experience sampling methods (Pejovic et al. 2016) and for engaging large groups of adults in cognitive measurement tasks (Steyvers and Benjamin 2019). Researchers on children's sleep have even leveraged data about infants' developing sleep cycles (Mindell et al. 2016). The next frontier is the use of app—based methods to allow reseachers (and parents) to collect data about children's language.

We are currently at work on an app called Wordful, which would allow parents to keep a running diary of their child's vocabulary. In our current prototype, the Wordful interface allows parents to transcribe specific words that their child has uttered. But, it also provides a novel interface that allows parents to enter words more quickly by swiping "cards" showing particular words, indicating whether their child produces the word on the card with a single manual response. The content of these cards can then be chosen adaptively to maximize the information content of a given amount of swipes. Unlike other adaptive testing methods, however, in the case of Wordful, parents can decide how many words they wish to enter and can come back again and again over weeks or months to update their estimates.

We have conducted a preliminary trial of Wordful in which we recruited parents through social media and invited them to fill out an online CDI using web—cdi (Meylan et al. 2019). We then encouraged them to use the Wordful app for 3—4 weeks, providing notifications periodically to draw them back into the app. At the end of the study period we gathered a further web—cdi administation. Our data (N=97) suggest reasonable correlations between web—cdi and Wordful data (r = .49 and .54 for the initial and final CDI administrations respectively), comparable to correlations between the two CDI administrations (.59) though lower than the longitudinal correlations reported in Chapter 4. Further, by virtue of its flexible data collection interface, Wordful is not limited to the inclusion of a single word list — instead, we were able to broaden the sample of words that parents were asked about and recover age of acquisition information for new words.

Mobile app—based methods like Wordful provide a platform for beginning to address some of the fundamental weakenesses of CDI data that we enumerated above. For example, while parents will never be trained phoneticians, they can be asked follow—up questions about the nature of their child's utterance that chould shed light on the phonological trajectory of specific word forms. They could even be encouraged to record and upload audio of specific wordforms of interest.

A mobile platform could be used to investigate semantic generalization issues as well. Imagine that parents might log a concept like "cat": immediately afterwards, parents could be asked to show their child pictures on the mobile device and ask "is this a cat." By varying the type of picture (e.g., canonical cat, non—canonical cat, dog, tiger, etc.), such an experiment could probe the child's generalizations at relevant times during the process of acquisition. While many empirical challenges remain to be solved to make this method practical, the combination of mobile experimentation and parent observation could be very powerful for probing word learning in the moment that it is happening.

Overall, we are hopeful about the potential role of technology in broadening the applicability of parent report to fundamental questions in language development. While no single approach is right for every question, we hope that efforts to broaden and deepen the CDI approach will yield insights into yet a wider range of facets of early language.

18.4 Conclusions

In addition to the specific methodological limitations of the instrument, our work here represents a potential upper bound on what can be done with data of this type. Even if we continue to accumulate contributions to Wordbank, we are unlikely to gain access in the near future to a deeper set of

longitudinal data or a far broader sample of languages. Future work on early vocabulary is likely to require new datasets, perhaps gathered by using the tools described above.

In addition, we feel hopeful that there are broader morals to our approach here that can be applied to other datasets. In particular, we hope the fundamental theme of our analyses — consistency and variability — can be applied more broadly. For any dataset like ours that includes samples of children distributed across groups, phenomena can be assessed in terms of their relative consistency across these groups. To the extent that these groups cross–cut important features of human experience such as culture, language, or national origin, consistencies across these features provide statistical evidence towards the project of finding candidate universals of the relevant domain. Will this approach be fruitful in revealing either universals of human development or major dimensions of variation in human experience? Only time will tell. We hope, however, that we have created a potential template to enable such investigations.