

ARTICLE

Documenting the acquisition of indigenous languages

Clifton PYE*

The University of Kansas

*Corresponding author: Clifton Pye, University of Kansas - Department of Linguistics, 1541 Lilac Lane, Lawrence, Kansas 66045, United States. Email: pyersqr@ku.edu

(Received 19 March 2019; revised 19 August 2019; accepted 7 May 2020; first published online 5 June 2020)

Abstract

The outstanding property of human language is its diversity, and yet acquisition data is only available for three percent of the world's 6000+ spoken languages. Due to the rapid pace of language loss, it may not be possible to document how children acquire half of the world's indigenous languages in as little as two decades. This loss permanently diminishes the scope of acquisition theory by removing its empirical base. In the face of pervasive language loss, the question of how best to document the language of the last children to acquire indigenous languages assumes critical importance. A collaborative effort by researchers is required to identify the most efficient procedures for documenting children's language, and share them worldwide. This paper makes the case for documenting diversity and outlines steps needed to accomplish this goal.

Keywords: documentation; indigenous languages; acquisition toolkit

Introduction

The imminent loss of fifty percent of human languages by the end of the century (Crystal, 2000; Grenoble & Whaley, 1998; Hale, 1992; Krauss, 1992) will irrevocably impact empirical research on language acquisition. Currently research on language acquisition addresses many specialized topics across several hundred languages constituting roughly three percent of the world's 6000+ spoken languages (Crystal, 2014). This research has not resulted in a comprehensive description of language acquisition for even this small set of languages (Berman, 2014). We lack comparative data on children's early vocabulary, children's early sounds, children's verbs, children's use of negation, and children's use of topics for more than a handful of languages. The data needed to construct a representative sample of languages to test acquisition theories in a meaningful way do not exist.

Researchers have previously noted the paucity of documentation on the acquisition of the world's languages (Ambridge & Lieven, 2011; Berman, 2014; Lieven & Stoll, 2013; Lust, 2006; Pye, 2017; Slobin, 1985; Stoll, 2016). Crystal (2014), for example, observed that over 40 years the *Journal of Child Language* only published papers on the

© The Author(s), 2020. Published by Cambridge University Press

acquisition of 20 languages spoken outside Europe. He added that “Given the presence of 6,000+ languages in the world, it seems we have still some way to go to put typological flesh on our hypotheses, with several language families having no representation at all in this.” Kelly, Forshaw, Nordlinger and Wigglesworth (2015: 286) observe that “There is a dawning realization that the field of child language needs data from the broadest typological array of languages and language-learning environments.”

Far more serious are the linguistic features that are missing from the acquisition record. We do not know what unique features children produce in the majority of languages that remain undocumented. Textbooks on language acquisition do not include information about the acquisition noun incorporation, split ergativity, switch reference, obviation or ballistic stress. The urgent need to document the acquisition of children’s speech in 6,000+ languages can only be addressed by reorienting the field from an intensive focus on the acquisition of a few languages to the documentation of child language in thousands of languages. This reorientation sets new theoretical goals, transforms our understanding of what children acquire, and brings immense challenges. My goal in this paper is to make the case for reorienting the field in such a drastic manner, and outline the steps needed to accomplish it.

I focus on documenting the acquisition of indigenous languages in this article because these languages are critically endangered. As stated on the United Nations webpage <<https://www.un.org/development/desa/dspd/2019/01/2019-international-year-of-indigenous-languages/>> “In 2016, the United Nations General Assembly adopted a resolution (A/RES/71/178) proclaiming 2019 as the International Year of Indigenous Languages, based on a recommendation by the Permanent Forum on Indigenous Issues. At the time, the Permanent Forum expressed concern that 40 per cent of the world’s estimated 6,700 languages were in danger of disappearing – the majority belonging to indigenous peoples.”

Indigenous languages are the dark matter of the linguistic universe. According to Ethnologue <<https://www.ethnologue.com/statistics/size>> just 19.5% of the world’s languages have 100,000 or more speakers. Assuming that language endangerment is inversely correlated with the number of speakers suggests that it is critical to document how children acquire the 80.5% of the world’s languages with fewer than 100,000 speakers. These languages possess 80.5% of the typological variation in human language that is needed to fully understand children’s ability to acquire language. Testing a theory of language acquisition with data from 19.5% of human languages cannot be justified. A sample that is typologically skewed toward Indo-European languages makes a bad situation worse.

The implications of language loss for language acquisition research

The Sports Sunday section of the June 22, 2014 New York Times ran a picture on its first page labeled “Indigenous Sateré residents of Nova Belo Horizonte, an impoverished village in the Amazon basin, watched the World Cup match between Brazil and Mexico on Tuesday.” The picture accompanies a story by Jeré Longman with the title “The Everywhere Game: The World Cup’s Fevered Grip Can Be Felt Even in the Deep, Isolated Amazon Rain Forest, Where Outsiders Seldom Visit.” The picture shows a family group sitting in a thatched-roof house watching the game on television. Three men and a 10-year-old girl sit in the front row, while the women and smaller children sit in the back. While the reporter marveled at the natives’ determination to watch the game, a linguist sees a picture of the Sateré

language being replaced by Brazilian Portuguese. The picture captures a two-hour period in which the younger children observe how the global economy claims the linguistic affiliation of the dominant family members.

Another New York Times story “A Long-Frozen Place, Melting and Changing Right Under Their Feet” by Neil MacFarquhar (August 4, 2019) describes how global warming is reshaping the arctic region of Yakutia in Russia. The January temperature there now reaches minus 50 Fahrenheit rather than minus 75, which is enough of a difference to melt the permafrost in unpredictable ways and make travel more difficult. MacFarquhar writes that “Indigenous peoples are more threatened than ever ... leaving them feeling baffled, unsettled, helpless, depressed and irritated.” Climate change will contribute to the cultural extinction of many traditional societies and their languages.

Today, predictions of language loss are nearly as ubiquitous as warnings of global climate change. Indeed, rising sea levels and desertification will severely impact marginal societies who typically speak endangered languages. Political conflict, global pandemics and criminal cartels adversely impact marginal communities that have maintained indigenous languages for millennia. Language loss also results from government aid to communities in the form of rural electrification, roads, community schools and cellphone towers, which improve the lives of many villagers at the cost of introducing the majority language to minority communities. I still marvel at watching nature documentaries and professional wrestling matches in adobe homes in Guatemala that lacked indoor plumbing.

The contexts for language loss have been widely documented (cf. Brenzinger, 2007; Campbell, 2017; Crystal, 2000; Dorian, 2004; Grenoble & Whaley, 1998; Thomason, 2015). What is not appreciated so well is the increasing rate of language loss and its profound implications for research on language acquisition. The linguistic consensus is that half of all languages spoken today will be lost by the end of this century (Campbell & Belew, 2018; Crystal, 2000; Grenoble & Whaley, 1998; Hale, 1992; Krauss, 1992). Languages become extinct when their last speakers die. Assuming that these last speakers are between the ages of 60 and 80 years old when they die implies that they will be born sometime between 2020 and 2040. In other words, only two decades remain in which to document how children acquire over half of the world’s languages.

The documentation of indigenous languages of Mexico is instructive. McQuown (1955: 501) claimed that “in one small portion of the area, in Mexico just north of the Isthmus of Tehuantepec, one finds a diversity of linguistic type hard to match on an entire continent in the Old World.” Ethnologue (Eberhard, Simons & Fennig, 2019) estimates there are 282 indigenous languages in Mexico (<https://www.ethnologue.com/country/MX>, accessed 3/3/19). Campbell and Belew (2018) catalogue the endangered languages in Mexico as well as in Central America and the Caribbean. Table 1 lists ten indigenous languages in Mexico where some level of documentation for language acquisition exists, with citations for some of this research (cf. Pfeiler, 2007). Acquisition studies of a few other indigenous languages in Mexico are available in student theses and dissertations that are not readily accessible. The 10 languages belong to four language families and constitute 3.5% of the 282 living indigenous languages in Mexico. Theories of language acquisition do not fully address these findings.

It has taken approximately 40 years to assemble this acquisition record for Mexico. At this pace, one language every four years, it would take another 1088 years to

Table 1. Published acquisition research for indigenous languages in Mexico

Family	Language	Acquisition Studies
Mayan	Ch'ol	de León, 2007; Pye, 2017; Pye, Pfeiler & Mateo Pedro, 2017b
	Tzeltal	Stross, 1969; Brown, 1994, 1998, 2001; Brown <i>et al.</i> , 2013
	Tzotzil	de León, 2001; Pye <i>et al.</i> , 2007; Brown <i>et al.</i> , 2013
	Wastec	Larsen, 1949; Pfeiler & Pye, 2012; Pye & Pfeiler, 2017a
	Yucatec	Carrillo Carreón, 2007; Pfeiler, 2002, 2003; Straight, 1976
Otomanguean	Northern Pame	Pye, Berthiaume & Pfeiler, 2020, in press
	Zapotec	de López, 2007; Stemberger & Chávez-Peón, 2014
Totonacan	Totonac	Montes Castañeda, 2014
Uto-Aztecan	Huichol	Gómez López, 1998, 2007
	Nahuatl	Ramírez Nava, 2005

document how children acquire the other 272 indigenous languages of Mexico. At twice this pace, it would still take 544 years. I doubt if we have another 50 years to accomplish this work. The scarcity of documentation is much the same for Africa, Australia, Indonesia, New Guinea and every other continent outside of Europe.

Documenting the acquisition of indigenous languages while there are still children who are acquiring them requires a fundamental shift in the conceptual framework for doing language acquisition research. There is an urgent need to develop an approach to document the acquisition of indigenous languages as rapidly as possible, archiving the research and returning the products of this research to the indigenous communities. Current methods for language acquisition research are not equal to the task of documenting the acquisition of four thousand languages in even eighty years (50 new languages each year for the rest of this century). The prospect of imminent language loss imposes a moral and scientific imperative to shift from intensive research on the acquisition of a few languages to comprehensively documenting the acquisition of the many endangered languages while the languages are still being acquired by children.

The importance of documenting the acquisition of indigenous languages

The importance of documenting the acquisition of indigenous languages is still not widely appreciated (Hale, 1998; Mithun, 1998; Pye, 2017). Indigenous languages have grammatical features that do not exist in better studied languages (Evans & Levinson, 2009; Hale, 1992, 1998; Krause, 1992; Mithun, 1998). We lack information on the acquisition of ejective, implosive, double articulated and click consonants. Little information is available on the acquisition of laryngealized vowels and grammatical tone. Insufficient information is available on the acquisition of ergative and active agreement, and especially on the acquisition of languages with split ergativity (Bavin & Stoll, 2013; Pye, 1990; Pye, Pfeiler & Mateo Pedro, 2013). Also lacking is information on the acquisition of object-initial basic word orders, noun/oblique

incorporation and grammaticalized complex animacy hierarchies. Even though the acquisition of pronouns has been a longstanding focus of research, we know little about the acquisition of person marking across languages that use pronouns, pronominal clitics and nominal classifiers (Siewierska, 2004).

Evans and Levinson (2009: 446) observe that “The diversity of language is, from a biological point of view, its most remarkable property – there is no other animal whose communication system varies both in form and content. It presupposes an extraordinary plasticity and powerful learning abilities able to cope with variation at every level of the language system.” Haspelmath (2010: 85) adds that “Different languages represent historical accidents and (unless they influenced each other via language contact or derive from a common ancestor) the categories of one language have no causal connection to the categories of another language.”

Every linguistic feature, whether they are subjects, passives or stop consonants, is unique to each language (Comrie, 1981; Port & Leary, 2005; Stassen, 2011). The difference between the allophonic contrast between plain and aspirated stops in English, e.g., [p] vs. [p^h], and the phonemic contrast between similar sounds in Thai, e.g., /p/ vs. /p^h/, reflects the different ways in which the sounds are organized in the two languages (cf. Port & Leary, 2005). Evans and Levinson (2009: 433) note that “experts on sound systems are abandoning the Jakobsonian idea of a fixed set of parameters from which languages draw their phonological inventories, in favor of a model where languages can recruit their own sound systems from fine phonetic details that vary in almost unlimited ways ...”

The contrast between *put in* in English and *kkita* in Korean reflects the difference between the IN/ON semantic contrast in English and the tight/loose-fitting semantic contrast in Korean (Bowerman & Choi, 2001). Bowerman and Choi suggest that “early semantic development involves a pervasive interaction between nonlinguistic conceptual development and the semantic categories of the input language, not just a one-way mapping from preexisting concepts” (2001: 477). The world’s languages embody different constellations of linguistic features and provide the primary empirical evidence of children’s capacity to learn a myriad of language-specific features. The following three examples illustrate diverse linguistic features that remain outside the scope of acquisition theories.

Sumu is an endangered Misumalpan language spoken by 7,000 people in Honduras and Nicaragua. Most other Misumalpan languages have died out. Sumu has a switch-reference system that tracks reference between clauses and distinguishes the case of the same referent from that of a switched referent. Sentences (1a-b) illustrate the Sumu switch-reference system (Hale, 1998).

- (1) a. yang nawah tal-i îr-ikda.
 I tiger see-PROX run-PAST.1SG
 ‘On seeing the tiger, I ran away (same subject).’
 b. yang nawah tal-ing îr-ida.
 I tiger see-OBV.1SG run-PAST.3SG
 ‘Upon my seeing the tiger, it ran away (different subject).’

In (1a), the verb *tal* ‘see’ has the suffix *-i* to indicate that the subject has the same referent as the subject of the following clause. In (1b), the same verb has the suffix *-ing* to indicate that its subject has a referent that is not the same as the subject of the

following clause. Research on the acquisition of switch reference would supply essential information on children's ability to track referents across clauses.

In 1966 Ken Hale published a brief note on the pronominal system of Lardil, an Australian aboriginal language spoken today by fewer than 5 people on Mornington Island. Norvin Richards noted that Lardil "was deliberately destroyed" through a program of assimilation and relocation (Wright, 2003). Hale reported that Lardil pronouns mark a distinction between harmonic and disharmonic generations (2). The harmonic generations include kinship relations between parents and grandparents or between parents and grandchildren. The disharmonic generations include the kinship relations between children and parents. We are unlikely to ever know how children acquire the harmonic distinction in pronouns, and how they would extend the pronouns to new acquaintances.

(2) Lardil pronouns (Hale, 1966)

	Harmonic	Disharmonic
Second person dual	kirri	nyiinki
Third person dual	birri	nyiinki
Second person plural	kili	kilmu
Third person plural	bili	bilmu

Mam is an Eastern Mayan language with a half million speakers who live in western Guatemala (England, 1983). Mam generally uses an ergative set of verb affixes to cross-reference the subject of transitive verbs and an absolutive set of verb clitics to cross-reference the object of transitive verbs and the subject of intransitive verbs. Mam extends the ergative set of person markers to cross-reference the object of transitive verbs in temporal clauses (3). These extensions result in transitive verbs with two ergative markers, one cross-referencing the subject, and the other cross-referencing the object.

(3) Mam verb with two ergative person markers (England, 1983: 259)

ok	t-kuʔ-ʂ	ky-awa-ʔn	ʂxaal	kxoʔn ...
when	ERG3SG-DIR-DIR	ERG3PL-plant-DEPTV	person	cornfield
"When the people plant the cornfield ..."				

The third person singular ergative marker *t-* in (3) cross-references the cornfield, whereas the third person plural ergative marker *ky-* cross-references the people. Children acquiring Mam must learn when to use ergative affixes to cross-reference the objects of transitive verbs and when to use absolutive clitics. Fortunately, acquisition data are available for Mam and show that two-year-old children understand the form and function of extended ergativity (see below).

We cannot fully understand children's ability to acquire language until we have information on the diverse language features that children acquire, including switch-reference, harmonic pronouns and extended ergative marking. The only way to gain this understanding is to document how children acquire thousands of languages that instantiate the full panoply of grammatical features that occur in human languages. We cannot understand the theoretical implications of acquiring the unique features of languages without the primary data from children.

Impediments to research on indigenous languages

The paucity of research on the acquisition of indigenous languages is evidence of a systemic gap in the practice of acquisition research in the sense that researchers lack a framework that motivates the study of indigenous languages on a scale required to document the acquisition of these languages before they go extinct. The absence of research on indigenous language acquisition restricts understanding of the significance of such research, which restricts the resources devoted to such research.

One criticism of rapid documentation projects of the kind I advocate is that such research should not be undertaken before grammars and dictionaries become available for the adult language. Without basic linguistic information, the investigator will lack the requisite knowledge that is needed to understand how the language really works. My response to this criticism is twofold. First, we are still waiting for definitive grammars for most languages. We do not have the luxury of waiting for linguists to produce grammars before documenting the acquisition of endangered languages. Secondly, the recordings of children interacting with their families reveal a verbal repertoire that will be missed in standard linguistic fieldwork with adult informants. In any case, this criticism misses the point of documenting language acquisition so that a permanent record will exist when a better understanding of the adult language becomes available.

Another criticism is that children may no longer be acquiring indigenous languages in monolingual communities. Most indigenous languages have been in contact with national languages for centuries and have borrowed heavily from the national languages (Thomason & Kaufman, 1988). Children in these situations may produce a high percentage of words in the national language. Such situations appear to be problematic at first glance, but actually present more opportunities than problems (cf. Pye, 2013; Pye *et al.*, 2020). We do not have a good record of children's language for communities that are in the process of abandoning the indigenous language. Such information would prove vital to programs, which try to strengthen the use of indigenous languages. Additionally, speakers do not borrow words from the national language at random. They tend to borrow nouns for consumer goods and retain verbs from the indigenous language thereby preserving most of the grammar of the indigenous language. There is much we can learn about the acquisition of indigenous languages in mixed communities.

A reviewer asked why the focus should be on documenting indigenous languages with small populations (under 100,000 speakers) when it would be much easier to document the acquisition of languages with large populations and educated speakers. While it is much easier to research language acquisition with large populations, there is no rush to do so. It will be possible to document how children acquire these languages for the foreseeable future. Moreover, acquisition data are already available for most languages with more than 50 million speakers: Chinese, Spanish, English, Hindi, Arabic, Bengali, Portuguese, Russian, Japanese, Turkish, Korean, French, German and Italian. It will not be possible to document how children acquire endangered languages much longer.

The time for making excuses for not documenting how children acquire a majority of the world's languages has passed. Documenting the acquisition of the world's languages requires urgent attention now.

In the following sections I outline the infrastructure that is necessary to increase the documentation of endangered languages in this century. In the next section of the

paper, I propose a protocol for language acquisition research whose aim is to produce a description of language acquisition that is as comprehensive as possible given the limited time available to record children. The development of a standard protocol for language acquisition research should produce a dataset that accelerates cross-linguistic comparisons. The five studies that I propose document basic features of the acquisition process and can serve as a nucleus for future investigations. In the third section, I discuss the infrastructure that is necessary to promote the documentation of acquisition in more languages. A critical step in this endeavor will be to develop a coordinated plan for determining the languages that are most in need of documentation. The final section of the paper discusses the broader impacts that such studies will have on language documentation and language maintenance.

A framework for the rapid documentation of language acquisition

In this section I sketch a course of action for increasing the documentation of endangered languages in this century. The number of languages that need documentation and the abbreviated time that is available for documenting them impose strict limits on any documentation plan. I propose developing a toolkit for language acquisition research whose primary aim is to document language acquisition in a manner that is both efficient and comprehensive. This toolkit can be shared with anyone interested in documenting the acquisition of a language in order to shift the documentary burden from a small number of dedicated researchers to a distributed network of documentarians linked through the internet and local universities. Such a network will enable speakers of indigenous languages to assume much of the responsibility for documenting the acquisition of their own languages, and hopefully return the products of their efforts to their community.

Documenting children's language has five phases: 1) finding and recording the children and their families, 2) transcribing the recordings, 3) archiving the recordings and transcriptions, 4) analyzing the results and 5) returning the results to the community. It is essential to identify ways to make each phase as rapid and efficient as possible. The transcription and analysis phases generally take the most time to complete so anything that can be done to shorten these two phases will increase the number of languages that can be documented. I discuss each phase in turn.

Phase I: Recording children and their families

Many textbooks and field manuals offer excellent guidelines for language acquisition research. Brown (1973) provides the classic description for recording and analyzing longitudinal samples of children's language, and the field manual edited by Slobin (1967) provides guides to documenting different features of children's language. Textbooks by Ambridge and Lieven (2011), Clark (2003), Ingram (1989) and Lust (2006) describe more recent research procedures. Also worth consulting are guides to assessing children's language such as the one edited by Miller (1981). Ingram (1981, 1989) provides procedures for assessing children's phonological development. More recent publications identify the best practices for working with children and families in remote field conditions (cf. Demuth, 1996b; Eisenbeiss, 2006; Kelly & Nordlinger, 2014; Kelly *et al.*, 2015; Stoll & Lieven, 2014).

Time will always be the limiting factor. I estimate that a one-hour recording generally requires two weeks to transcribe. This rate allows 24 hour-long recordings

to be transcribed by one person per year, allowing some time away for holidays and accidents. This schedule limits the documentation to recording one child twice a month for one hour over the course of one year, recording two children twice a month for a half-hour, or recording eight children for three hours each. Stoll (2016:145) suggests recording each child for up to five hours in a single week, but such a schedule may not be practical in all cultures. Of course, more transcribers or more time allows more recordings to be transcribed, but increases costs. The recording schedule is highly dependent on the number and ages of the available children, which differ by the size of the language community. Recording three two-year-old children over the course of a year is ideal (cf. Brown, 1973), but cross-sectional recordings of five children aged 2;0, 2;6, 3;0, 3;6 and 4;0 also yield useful information (cf. Cook, 2006; Mithun, 1989). Milroy (1987) provides a classic discussion of the factors that influence the quality of language samples.

Language acquisition researchers will need to seek permission from the families and language community to conduct their research. This is a delicate process that is difficult to do without a deep understanding of the community and its language situation. A very real hazard in work with indigenous communities is that an outside investigator will unknowingly stumble into a situation that leads to malign rumors about the purposes of the research. Many indigenous communities have good reason to be suspicious of outsiders because of their experiences with strangers taking their resources or their children (Nolan, 2020). The process can be made easier by seeking the help of experts who work with the community, who can identify the conditions for doing research in the community, who can introduce the researcher to community members who can aid in the research, and who can vouch for the integrity of the researcher. These intermediaries can be linguists or health workers or ideally members of the community themselves such as college students or school teachers.

Working with small children will likely attract attention from the community as well as institutional review boards. It is essential to provide children and their families with information about the purpose of language documentation, and how their identities will be protected. They should also be informed about the significance of their contributions. It is of utmost importance to protect the children and their families from harm so it may be necessary to use audio-recorders rather than video cameras in order to protect family identities while making the recordings available on a community language archive. This is not a problem that can be taken lightly, and must be tailored to the needs of the individual families and communities.

Phase II: Transcription

The pressure of time severely constrains the level of transcription that is useful in rapid documentation. I advocate the use of a minimalist transcription format that transcribes what the children say, the adult interpretation of the children's utterances, and a translation of the adult interpretation into the national language, cf. <<http://pysqr.org/minimal/>>. An orthographic system is an enormous help to the transcription process, but otherwise a broad phonetic transcription will serve. Additional annotations for morphological or discourse analysis complicate transcription and introduce errors. It is best to avoid adding grammatical codes in order to maintain a strict separation between the transcription and analysis phases of the project.

One broader impact of language acquisition research should be the training of members of the language community to transcribe and analyze their own language.

It is hard to overestimate the impact that teaching speakers to write their own language on computers has in terms of elevating their respect for their language. Adding speakers from the language community to the research team, asking them to record, transcribe and analyze the data they record, and looking for ways to return written products of the research to the community help to elevate respect for the indigenous language in the community. It also promotes understanding in the community for the acquisition research, and extends the training of indigenous peoples to document their own languages (England, 2003, 2007; Hinton, 2001; Gehr, 2013). One way to return written documents to the community is to make books for the participating families from the transcriptions of their children's recordings that show how their language is written. Adding photographs of the family members to the books adds to their impact.

The employment of native speakers in the transcription phase of the project constrains the level of transcription. Adding morpheme glosses and grammatical codes to the transcription makes it more difficult to recruit native speakers for the transcription work. Grammatical codes can be added after the basic transcriptions become available at a time when the investigator has a better understanding of the language. It is far more important to make use of the linguistic and cultural insights that native speakers bring to the interpretation of children's utterances.

Phase III: Archiving the recordings and transcriptions

A plan for archiving the recordings and transcriptions should be a central part of any language documentation project. The CHILDES archive (MacWhinney & Snow, 1985) is well-known in language acquisition research, but many other language archives now exist such as The Archive of the Indigenous Languages of Latin America (<https://ailla.utexas.org>), and the Documenting Endangered Languages Programme <<https://www.eldp.net>> at the School of Oriental and African Languages. The Language in Time and Space webpage supplies a list of language archives around the world (<https://lucian.uchicago.edu/blogs/langtimespace/resources/resourcesarchives/>). The Open Language Archives Community offers information on the best current practice for the digital archiving of language resources (<http://www.language-archives.org/>).

Language acquisition data should be archived with future centuries in mind so it is important to consider how changes in technology may render the collection inaccessible. The material will be most useful if it is accessible in the cloud, but the internet brings a host of problems that include hacking and possible misappropriation of the material. The safety of the children and the community has to be paramount when archiving the recordings.

Archiving the recordings generates considerable costs. I estimate that storing a one-hour, high-definition, video recording with its transcription requires 5 gigabits. Twenty-four recordings for one child would require approximately 100 gigabits to store. It is reasonable to assume that the recordings for one language would require 500 gigabits, and the recordings for a thousand languages would amount to 500,000 gigabits.

Recordings of child language document the culmination of the last two-hundred thousand years of linguistic evolution and constitute a critical part of the world's intellectual heritage. The creation and protection of language archives requires a global initiative similar in scope to the world's natural history museums. Funding sources for a project of this magnitude would have to include a mix of public and private sources. Publicity for the project would increase its visibility and help secure funding.

Phase IV: Analysis

A shift to a model of archiving acquisition data for the world's indigenous languages in two decades is the main reason to abandon the model of the individual investigator who collects and analyzes child data from a single language over a period of several decades. Archived acquisition data for indigenous languages will retain their significance long after their collectors have retired and the acquisition theories of the day are forgotten. Shifting the orientation of acquisition research to documentation places a premium on devising efficient analytical methods and collaborative modes of research.

Researchers must always investigate the grammar of an indigenous language on its own terms (Boas, 1911). In creating a protocol for child language documentation, it is essential to invent analysis techniques that can be used with languages of all types and that use minimal transcriptions. One way to do this is to focus attention on analyzing the words that children produce. Lexical acquisition stands at the heart of acquisition research, and only requires minimal assumptions about the adult grammar. It is also easy to do with minimal transcriptions.

The first step in lexical analysis is to produce a lexical concordance for the children's words. A concordance provides the utterance contexts for a child's production of each word in the recording. It is best to use the adult targets as the concordance keys in order to group together a child's attempts at each adult word. The first publication for any investigation of children's language should be a lexical concordance for each recording. A portion of a lexical concordance for a two-year-old child acquiring Northern Pame is shown in (4).

(4) A lexical concordance for a two-year-old Northern Pame speaker

Word	Begin Time	Child's Utterance	Adult Target	Translation
daʔtsəlʔ	03:54.9	kiʔil	daʔtsəlʔ	It bit.
daʔuap	20:56.8	ane	daʔuap	It threw it.
danàs	12:44.2	nas	danàs	orange
dapaj	05:15.5	paj	dapaj	tomato
	16:26.2	paj	dapaj	tomato
	16:30.6	paj	dapaj	tomato
	07:21.6	pa	dapaj	tomato

The lexical concordance in (4) shows the variation in the child's production of four lexical types. The time codes make it possible to locate the child's utterances in the original recording. The concordance provides basic data on the lexical categories that the child produced as well as their type and token frequencies. Information on concordance programs is posted on my webpage <<http://pyersqr.org/minimal/>>. In the remainder of this section I will sketch five studies of children's lexicons that can be done with a lexical concordance.

Study 1: The Lexicon

The lexicon has been a neglected area of research on language acquisition partly due to a lack of appreciation for the degree of lexical variation across languages. Typologists have long emphasized the differences between words across languages (Haspelmath, 2011; van Gijn & Zúñiga, 2014). One source of variation is the difference between word units defined by prosody and word units defined by the grammar (Bickel & Nichols, 2007; Mithun, 2014).

The grammatical classes of words differ between languages and therefore define a basic research question that concerns how children establish the grammatical classes for words in the adult language. Languages may have distinct noun classes for alienable and inalienable nouns, grammatical gender or arbitrary noun classes. Languages may have distinct verb classes for transitive, intransitive, stative, existential and derived verb classes or arbitrary classes for verbs (Grinevald, 2003). Languages may or may not have distinct classes of adjectives, articles, pronouns, prepositions or positional verbs. The particle classes never cease to puzzle linguists. Publishing a simple record of children's first words is already a major step in documenting the acquisition of a language (cf. Nelson, 1973, 2014; Rescorla, 1980).

Adding codes for the lexical categories to the lexical concordance makes it possible to sort children's words by their lexical category. The sorting can be accomplished rapidly and accurately in any spreadsheet program. The example in (5) adds a column for lexical categories to the concordance in (4). The code LD refers to one of four verb categories specific to Northern Pame (Berthiaume, 2012) while the code CN is used for common nouns.

(5) A lexical concordance for a Northern Pame child with lexical category codes

Word	Category	Begin Time	Child's Utterance	Adult Target	Translation
daʔtsəlʔ	LD	03:54.9	kiʔil	daʔtsəlʔ	It bit.
daʔuaŋ	LD	20:56.8	ane	daʔuaŋ	It threw it.
danās	CN	12:44.2	nas	danās	orange
dapaj	CN	05:15.5	paj	dapaj	tomato
		16:26.2	paj	dapaj	tomato
		16:30.6	paj	dapaj	tomato
		07:21.6	pa	dapaj	tomato

Surprisingly little information is available on children's lexical development for indigenous languages. The Northern Pame example shows a child's production of a language-specific category of verbs. The examples in (6) show children speaking the Mayan language Mam using intransitive verbs in contexts where English or Spanish speakers would use transitive verbs.

(6) Mam children's use of intransitive motion verbs as substitutes for transitive verbs (Pye, 2017)

a. WEN (2;0.2)

kuʔ pe tuʔnʔ

= ma pa 0-kub' kape t-uʔn-a

REC already ABS3-go.down coffee ERG2-by-ENC

'Did you already pick the coffee?' (lit. 'Did the coffee already go down by you?')

b. JOS (2;7)

maʔ af x wonn.

= ma xaw-ʂ xal w-uʔn-a.

REC rise-away CL.NH ERG1-by-ENC.

'I lifted it up.' (lit. 'It rose up by me.')

WEN produced the oblique agent phrase using the relational noun *t-uʔn-a* 'by you' in (6a), and JOS produced the oblique agent phrase *w-uʔn-a* 'by me' in (6b). Relational

nouns take the same possessive morphology as common nouns, but are used to express syntactic relations. The use of relational nouns to express agents in oblique phrases enables Mayan languages to use intransitive verbs to express events that involve agents and patients. Grammatical words such as relational nouns are an important part of the lexical acquisition process and constitute their own domain of study. Grammatical words are also one of the primary sources of diversity across languages.

Study 2: Segmental inventories

Phonological research can be the most rewarding research to undertake with small language samples because the few, isolated utterances that children produce contain a lot of data on the sounds that children produce as well as the substitutions that children make for adult sounds. Phonological research provides some of the richest data on language variation in child and adult speech and we can use these variable realizations to gain insight into the developing stability of children's grammar. An analysis of child phonologies in six Mayan languages demonstrated that the individual variation for children acquiring each language differed from the variation for children acquiring the other languages (Pye, Pfeiler, Mateo Pedro & Stengel, 2017c).

A basic phonological study consists of providing an inventory of the sounds that children produce in their words. David Ingram (1981, 1989) provides an efficient method for identifying children's basic segments from the noise of their background variation. Phonological studies begin with lists of words that children produce so a lexical concordance feeds directly into studies of phonological development.

Ingram's method is based on selecting a representative phonetic type from the variety of phonetic tokens children produce for individual words. The representative phonetic type serves as a compromise between all of the words children produce and their variations on each word. The Northern Pame example in (5) shows four attempts by the child to say the word *dapaj* 'tomato'. The child produced the word as [paj] in three of his four attempts so [paj] would be selected as the representative phonetic type for the adult word *dapaj*. Once the list of representative phonetic types is constructed, the investigator can use segment frequency across word types and a substitution analysis to construct an inventory of consonants and vowels that a child uses in initial, medial and final positions in words.

Incredibly, we lack consonant and vowel inventories for most of the languages where language acquisition research exists. Establishing standard procedures for extracting such data is essential. Table 2 shows the initial consonant inventories for children acquiring English, K'iche', Mam, Teenek and Chipewyan that were produced using Ingram's procedure. These data show that the children generally produced the nasal and plain stops as well as glides in initial position. They produce voiced stops if the adult language has them and not otherwise. They produce different types of affricates and fricatives depending on the phonological structure of the adult language. The children generally omit uvular and ejective stops. The children acquiring the three Mayan languages have initial consonant inventories that are more similar to one another than to the consonant inventories of children acquiring English and Chipewyan.

Segmental inventories like those in Table 2 are just a starting point for further study (cf. Menn, 1983; Bernhardt & Stemberger, 1998; Pye *et al.*, 2017c). These are only inventories of the consonants that the children produce in initial position so they should be augmented with inventories for consonants in medial and final positions as well as vowels. Each sound can be broken into a set of features for voice, place

Table 2. Initial consonant inventories for children acquiring English, K'iche', Mam, Teenek and Chipewyan

Language	Source	Initial consonant inventory							
English	Ingram, 1981	(m)	n	p	t	k	b d (g)	(f) (s) h	w
K'iche'	Pye <i>et al.</i> , 1987	(m)	n	p	t	k	ʔ* tʃ* (b)	x*	l* w*
Mam	Pye <i>et al.</i> , 2017c	m	n*	p	t*	k*	tʃ*		l w j
Teenek	Pye <i>et al.</i> , 2017c	m	n	p*	t*	k	tʃ b	(t') (k') θ ʃ h*	l (w)
Chipewyan	Cook, 2006	m	n	t	k	ʔ ts tʃ	b d g dz dʒ	(ts')	h (r) l (w) j

and manner of articulation. Studies of the children's feature inventories produce more information about the structure of children's phonological representations (cf. Pye *et al.*, 2017c; Rice & Avery, 1994).

Study 3: Prosody

Prosody has significant effects on the shapes of children's early words and yet remains one of the least studied aspects of child language acquisition. Prosodic effects are evident in the syllables that children produce and those that they omit. Prosody has the hierarchical structure shown in (7) where a phonological word contains at least one foot, the foot contains at least one syllable, and the syllable contains at least one mora (Hayes, 1995; Lahiri, 2001; Selkirk, 1984).

(7) Prosodic Hierarchy

PW	(Phonological Word)	<i>bananas</i>
Ft	(Foot)	<i>nanas</i>
σ	(Syllable)	<i>nas</i>
μ	(Mora)	<i>na</i>

Theories of prosodic development predict that children initially construct Minimal Words that contain a binary foot composed of either two mora or two syllables (Archibald, 1995; Demuth, 1996a, 2001; Fee, 1995; Fikkert, 1994). Demuth (1996a, p.181) states that “the evidence presented here points strongly to the fact that children's early words are not randomly truncated forms, but well-organized Minimal Words.”

A key part of assessing prosodic effects is understanding the nature of stress and how it is placed in different languages. Languages can assign primary stress to syllables by their weight, e.g., whether the syllable is open or closed, or by position, or both. Stress can manifest as a change in pitch in pitch-accent languages or by changes in vowel duration or amplitude. And then there is the phenomenon of ballistic stress, which DiCanio and Bennett (*in press*) describe as including “some/all of the following phonological characteristics: (1) fortis-initial onsets, (2) shorter vowel duration, (3) an abrupt, final drop in intensity, (4) tonal variation (specifically F_0 raising), (5) post-vocalic aspiration, and/or (6) coda devoicing.”

Prosodic effects are evident in the Northern Pame lexical concordance shown above in (5). Most words in Northern Pame have an initial syllable that is extrametrical in that it does not bear tone (Berthiaume, 2019). Children acquiring Northern Pame tend to omit the extrametrical syllables (Pye, Berthiaume & Pfeiler, *in press*). The Northern Pame child omits the initial syllable of the four words shown in (5), for example, producing the verb *daʔuaŋ* ‘It threw it’ as [ane].

An example of prosodic effects at the phrasal level is seen in the production of directional clitics in the Mayan languages K’iche’ and Mam. K’iche’ and Mam incorporate directional clitics into the verb complex in order to specify the direction of the agent’s movement in the event. The resulting verb complex has one marker for aspect and absolutive person marking. The examples in (8) show verb complexes with directional clitics in K’iche’ and Mam.

(8) Verb complexes with incorporated motion verbs

a. K’iche’

ʃ-ux-ee war-oq.
 CMP-ABS4-go sleep-DEP_{IV}
 ‘We went to sleep.’

b. Mam (England 1983:212)

ma 0-tsax t-tʃju-ʔn tʃeep tʃit.
 REC ABS3-come ERG3-grab-DEP_{TV} José bird
 ‘José came and grabbed the bird.’

These examples show that the K’iche’ directional *-ee* ‘go’ and the Mam directional *-tsax* ‘come’ are placed after the aspect and absolutive markers and before the verb, if intransitive (8a), or before the ergative subject marker if the verb is transitive (8b). The single absolutive person marker */-ux/* ‘we’ in (8a) cross-references the subject of the intransitive verb *war* ‘sleep’. The null third person absolutive marker in (8b) cross-references the object of the transitive verb *tʃju* ‘grab’.

We find a major difference in the acquisition of directionals in K’iche’ and Mam (Pye & Pfeiler, 2019). The use of directionals is largely absent in the speech of K’iche’ children until the age of three years, whereas children acquiring Mam use directionals before the age of two years. K’iche’ integrated the pieces of its verb complex while increasing the prominence of the final syllable of the complex. The pieces of the verb complex remained less prosodically integrated in Mam with the result that the directional clitics retained metrical prominence unlike the case in K’iche’. Two-year-old children acquiring K’iche’ and Mam respond to their structural features by producing the prosodically prominent syllables in the verb complex.

Study 4: Morphosyntax

Morphosyntax constitutes another undocumented area of acquisition research due to the paucity of information on the acquisition of indigenous languages. The unique characteristics of grammatical morphology in individual languages make it a challenge to generalize results (Pye, 2017; Stassen, 2011). Grammatical morphemes surface as prefixes, infixes, suffixes, circumfixes, clitics and reduplicated syllables, and may appear as stressed or unstressed syllables. Grammatical morphemes express a world of specialized meanings such as whether an event is out of control, whether the event was directly observed by the speaker, or whether the word occurs at the end of an intonational phrase (Pye, 1983).

The example of extended ergativity in the Mayan language Mam that I presented in (3) offers one example of how to use a lexical concordance for morphosyntactic analysis. Mam extends its ergative cross-reference markers to the subjects of intransitive verbs and the objects of transitive verbs in a limited number of adverbial phrases (England, 1983). Recordings of children acquiring Mam showed that contexts of extended ergativity in the children's speech occurred almost exclusively following the purpose adverb *ii* 'so that' (Pye, 2017). A portion of a two-year-old Mam child's lexical concordance entry for the purpose adverb is shown in (9). This example does not have morpheme glosses in keeping with the minimal transcription practice.

(9) Lexical concordance entry for the purpose adverb *ii* 'so that' for a Mam two-year-old

Word	Child's Utterance	Adult Target	Translation
<i>ii</i>	taaf x ^h unt kuun.	<i>ii</i> taxa xunt qu?n.	So that we turn another page.
	tax ^h tuj.	<i>ii</i> ttsax ntʃuuja.	So that my mother comes.
	taa ?unt.	<i>ii</i> taxa xunt xa.	So that another turns.
	taa ?unt ku?n.	<i>ii</i> taxa xunt xa qu?n.	So that we turn another.
	top x ^h luu.	<i>ii</i> tops xa tzaluu.	So that it's tied up there.
	top x ^h	<i>ii</i> tops xa.	So that it's tied up.
	kuux ^h kin.	<i>ii</i> tkub' xal tfikin.	So that it remains there.
	kuux ^h kin.	<i>ii</i> tkub' xal tfikin.	So that it remains there.

The lexical concordance example in (9) shows the child's productions that Mam speakers interpreted as purpose clauses even though the child systematically omitted the purpose adverb *ii*. The Mam speakers relied upon the discourse context and the child's use of extended ergative cross-referencing to interpret the child's utterances. The child produced the third person ergative marker /t-/ with the verbs *axs* 'turn' and *ops* 'tied'. She omitted the ergative prefix with the verbs *tsax* 'come' and *kub* 'remain' resulting in a 63% (5/8) frequency of use of extended ergative marking in this session. We could also examine the concordance entries for each of these verbs to see how often the child produced each verb in a context of extended ergativity. The lexical concordance makes such analyses easy without adding elaborate grammatical codes to the transcription.

Study 5: Verb arguments

Predicates and their arguments are the basic constituents of sentences. The production or omission of verb arguments in children's language has long been a topic of interest in the pro-drop literature (Hamann, 2002; Hyams, 2011). Beyond the simple production or omission of arguments further information can be learned about children's management of discourse by examining the types of arguments children produce. Languages express verb arguments as noun phrases, pronouns, noun classifiers, pronominal clitics, cross-reference markers on verbs, or some combination of all of these features (Siewierska, 2004).

A simple catalogue of the forms that children use in argument realization serves as an interesting benchmark of their knowledge of the discourse constraints on the use of arguments. Children must attend to the discourse enough to realize when they have to introduce a new topic into the conversation or refer to a previously mentioned entity. Once again, a lexical concordance is a valuable aid in analyzing the verb arguments that

children produce. Once the words have been coded for their lexical category, the verbs can be examined for their contexts of use. The example in (10) shows a Mam two-year-old's concordance entry for the intransitive verb *skueelan* 'to school'. This verb was borrowed from the Spanish noun *escuela* 'school', but Mam converts the word to an intransitive verb by adding the antipassive suffix /-an/.

(10) Lexical concordance entry for the Mam verb *skueelan* 'to school'

Word	Child's Utterance	Adult Target	Translation
skueelan	ke if. wej x ^h .	nchi skueelan bif. nchi skueelan xa?	The girls are studying in school. Are they studying in school?

Mam is interesting in part because it uses a set of noun classifiers in place of pronouns. The child's first utterance contains the noun classifier *bif*, which refers to familiar people namely the child's older siblings. The child used this classifier as the subject of the intransitive verb. The child's second use of the verb does not have an overt subject. The particle *xa* is a question marker. Both of the child's productions omit the progressive prefix /n-/ , the third person plural absolutive marker /chi-/ and the antipassive suffix /-an/.

The number and types of overt arguments that children produce as subjects and objects of verbs provide a clear indication of their acquisition of the discourse requirements of the adult language. Languages typically balance verb argument realization with the use of noun phrases, pronouns and agreement inflections on verbs. Linguists have noted that adult speakers of Mayan languages introduce new participants into the discourse as either direct objects of transitive verbs or subjects of intransitive verbs (Du Bois, 1987; England & Martin, 2003). Children acquiring the Mayan languages K'iche', Mam and Ch'ol produce few overt arguments as either subjects or objects, but, in many cases, their frequency of use of overt subjects with transitive verbs is infrequent relative to their use of overt subjects with intransitive verbs, which is similar to their frequency of use of direct objects (Table 3).

I used the label 'Proform' in Table 3 as a cover term for personal pronouns in K'iche' and Ch'ol, and for noun classifiers in Mam, which Mam uses in place of personal pronouns (cf. example 10 above). The K'iche' and Mam children realized verb arguments as proforms more frequently than the Ch'ol children. We need more information about how children manage argument realization, especially if they resort to the use of overt arguments instead of inflecting verbs for agreement (cf. Hamann, 2002).

When reviewing transcriptions of children speaking an indigenous language one constantly encounters small details of genuine theoretical significance. Two-year-olds produce verb contrasts between falling people and falling fruit (Northern Pame); they produce a demonstrative for something that can be heard, but not seen (Ch'ol); they produce a derivational suffix for sudden events (Wastec); they produce portmanteaux verb suffixes that mark a combination of verb transitivity and derivational status (K'iche'); they produce oblique agent phrases with intransitive verbs (Mam); and they omit singular prefixes on nouns (Northern Pame). Placing a greater emphasis on documentation than analysis will not stop investigators from publishing reports on the acquisition of language. It will expand awareness of all of the ways that languages have invented to say things and the things that two-year-olds say.

Table 3. Overt argument production by children in three Mayan languages (Pye, 2017).

Language	Child	Age	TV Subject		TV Object		IV Subject	
			NP n (%)	Proform n (%)	NP n (%)	Proform n (%)	NP n (%)	Proform n (%)
K'iche'	LIN	2;0		1 (6%)	7 (39%)		3 (13%)	2 (8%)
	TIY	2;1	1 (3%)	1 (3%)	3 (8%)	2 (5%)	1 (3%)	3 (9%)
Mam	WEN	2;0		5 (7%)	6 (8%)	2 (3%)	3 (4%)	4 (6%)
	JOS	2;6		1 (3%)	18 (49%)	6 (16%)	24 (28%)	10 (12%)
Chol	EMA	2;2	2 (2%)		6 (5%)	5 (4%)	4 (7%)	1 (2%)
	MAR	2;10	6 (9%)		12 (18%)	1 (2%)	4 (16%)	

Phase V: Returning the results to the language communities

The National Science Foundation in the United States requires investigators to address the broader impacts in research proposals. This requirement suggests that investigators go beyond simply seeking permission from the language community for their research and think of the ways in which their research can be returned to the language community and the wider world. The broader impact of research on language acquisition is an issue that we need to address systematically as a research community. Additional information on addressing the broader impact of linguistic research is available at the Documenting Endangered Languages Programme webpage <<https://www.eldp.net>> at the School of Oriental and African Languages (SOAS).

Many indigenous communities have bilingual teachers who help to ease the children's transition to kindergarten and first grade where the instruction occurs in the majority language. Bilingual teachers may be keenly interested in learning how to write their language and struggle to produce written materials in the indigenous language. Researchers can look for ways to forge alliances with teachers and community members in order to jointly develop materials for teaching reading and writing in the indigenous language.

An infrastructure for documenting the acquisition of indigenous languages

The need to document the acquisition of endangered languages as rapidly as possible demands a coordinated effort by researchers worldwide to map out the most efficient way to proceed. Investigators can post a set of maximally efficient procedures online, which can be used across research sites. Above all, there is an urgent need to identify the languages that are most in need of documentation so that scarce resources are not devoted to documenting languages that already receive substantial institutional support or are not in immediate danger (cf. Campbell & Belew, 2018 and UNESCO's Atlas of the World's Languages in Danger). Likewise, it is not possible to document the acquisition of languages that are no longer being acquired by children. The latter decision must be made with care because of the difficulty of locating the last few children acquiring a language. The sustainability of each language varies enormously depending on its local context. Languages have maintained themselves for centuries with populations of fewer than 100 speakers, but other languages with small populations vanish overnight once their children are sent away to school.

The effort to identify languages to study should proceed hand-in-hand with identifying local resources that can support language documentation. Selecting languages to document will depend upon a number of factors, including 1) whether two-year-old children are acquiring the language, 2) ease of access, 3) the political context, and 4) the available linguistic resources. The language acquisition researcher should make contact with linguists who have worked with the language community if only to access the grammars and dictionaries they have produced or know of. The linguist may be able to provide up-to-date information on the vitality of the language. In turn, the acquisition researcher can offer linguists information about child-directed speech, and child language, that will enhance a general understanding of the language.

If researchers cannot travel to the field site themselves it may be possible to ask linguists who are in the field to record children speaking the language. The language acquisition researcher must be cognizant of the limited time that linguists spend in the field, and so it may not be possible to make longitudinal recordings of the

children. Nevertheless, much can be learned from limited recordings of children at different ages. In this scenario, I suggest recording each child for at least one hour on three separate occasions. The recordings should be continuous in order to evaluate the overall rate of language production by the children and their caretakers during each recording session. Setting a standard length of time for recordings will make it easier to compare results across languages.

Working in small communities with limited resources makes it difficult to record many children. The documentation of indigenous language acquisition will likely produce small samples of children's speech. For this reason researchers will need to develop techniques for extracting the maximum amount of information from limited data sets. Cook (2006) and Mithun (1989) are examples of how investigators have made efficient use of limited data sets to document the acquisition of phonology and morphology respectively in two indigenous languages. Research with children acquiring majority languages can inform this endeavor as many studies of a few children acquiring these languages have been replicated in later studies with larger numbers of children (e.g., de Villiers & de Villiers, 1973). These replication studies provide an idea of what can be learned from a study of even a single child's speech.

The main consideration in research on the acquisition of endangered languages is the possibility that the investigator will only have a single opportunity to record acquisition data for posterity. If the investigator is fortunate enough to document the acquisition of an indigenous language they should assume their study will be the last to ever record children speaking the language (e.g., Kroeber, 1916; Nokony, 1978). In the absence of potential follow-up studies, the researcher must identify the most important features of the children's grammatical knowledge to record, including basic information on the children's lexicon, phonology, morphology, syntax, semantics and discourse. The imminent loss of indigenous languages demands that research on their acquisition be broad rather than deep. The acquisition data recorded today may well be used hundreds of years into the future to address theoretical issues that are as yet undreamed of.

Conclusion

The signature trait of human language is its wonderful diversity of forms and meanings. This diversity defines the core induction problem of language acquisition and deserves to be at the heart of acquisition research. Investigators cannot make valid claims about the process of language acquisition without some idea of how the process unfolds across thousands of languages. Many investigators have advocated adding studies of children acquiring more languages but to little effect. The next 20 years will be our last chance to document a major part of the diversity in child language. It is past time to take steps to increase awareness of language diversity and its significance for the field.

I propose an infrastructure for documenting the acquisition of indigenous languages that includes: 1) a toolkit for documenting language acquisition, 2) a standard set of analytical procedures, 3) a means to archive the results where they will be available to the language community and other researchers, and 4) a network for updating all of the above. The toolkit can be shared with linguists and members of language communities alike to promote wider documentation. Researchers can begin by publishing transcriptions and lexical concordances for every recording of child language they have at their disposal.

There is also a critical need for a better theoretical understanding of the implications that language diversity has for the problem of induction. An acquisition theory that does not address language diversity does not meet the basic test of observational adequacy. Fortunately, researchers can draw upon research on linguistic typology for information about diversity in human languages. The papers by Bowerman (2011), Evans and Levinson (2009), Haspelmath (2010), Port and Leary (2005) and Stassen (2011) provide a good starting point for understanding the diversity of human language. Online resources such as The World Atlas of Language Structures (WALS) provide basic information on the diversity of individual linguistic features (Dryer & Haspelmath, 2013).

Time is of the essence if we are to have any hope of documenting how children acquire a majority of the world's languages. To paraphrase Alfred Russel Wallace's (1863) remark on the significance of biological diversity: future ages will look back upon us as so immersed in the analysis of child language as to be blind to higher considerations. They will charge us with having culpably allowed the destruction of some of those records of diversity which we had it in our power to preserve; and while professing to regard every human language as worthy of study, yet, with a strange inconsistency, seeing many of them perish irrecoverably from the face of the earth, uncared for and unknown.

Acknowledgements. My colleagues on the Mayan Language Acquisition Project (Barbara Pfeiler, Penelope Brown, Lourdes de León and Pedro Mateo Pedro) have provided numerous insights into the workings of their languages, which increased my general understanding of Mayan language acquisition. Data collection for K'iche' was supported by grants from the Organization of American States and the Wenner Gren Foundation. Data collection for Ch'ol, Mam and Q'anjob'al was funded by grants from the National Science Foundation (BCS-0613120 and BCS-0515120) and the University of Kansas. Data collection for Wastec and Yucatec was funded by the Consejo Nacional de Ciencia y Tecnología de México (CB105596, CB4639-H, and CB27893-H) and from the Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica-Universidad Nacional Autónoma de México (IN401207) to Barbara Pfeiler. Data collection for Northern Pame was supported by a grant from the National Science Foundation (BCS-1360874).

All of these projects would not have been possible without the aid and support of the children's families as well as the efforts of the team of Wastec investigators: Leonarda Hernández Gutiérrez, Magdalena Martínez Enríquez, Alicia Hernández Martínez, and Andrés Cruz Cruz; the Yucatec investigators: Neifi Vermont Vermont, Andrés Dzib Dzib, and Adiel Mena Keb; the Ch'ol investigators: Pedro Gutiérrez Sánchez, Asunción López Pérez, and Melba del Carmen Martínez Pérez; the K'iche' investigators: Augustin Huix Huix, Pedro Quixtan Poz, Emilio Quij Huix and Santos Quij Huix; the Mam investigators: Ana Elizabeth López Ramirez, Juana Isabel López Morales, Sheny Ortíz García and Luis Hernandez López Ramirez and the Q'anjob'al investigators: Flora García, Diego Martínez Esteban, Francisco Pedro Mateo, Pedro Martínez Esteban, Efraín Ramón de León, Basilio Luin Bernabé, Basilio Sebastian Basilio. The Northern Pame investigators, Felix Baltazar Hernández and Constantino Gomez González, have made invaluable contributions to my latest project.

I would also like to thank the editor, Johanne Paradis, and associate editor, Cecile De Cat, as well as the two reviewers of this article for suggestions that have substantially improved the initial draft. I received many useful suggestions from audiences at the XX Encuentro de Adquisición del Lenguaje, Casa Chata, El Centro de Investigaciones y Estudios Superiores en Antropología Social, Mexico City, Mexico, May 2, 2019 and at the annual meeting of the Linguistics Society of America, January 4, 2020. I am responsible for any remaining errors.

Abbreviations

- 2 second person singular
- 3 third person singular

4	first person plural
ABS	absolute
CL	classifier
CMP	completive aspect
CN	common noun
COMP	complementizer
DEP	dependent status
DIR	directional clitic
ENC	enclitic
ERG	ergative
IV	intransitive verb
LD	LD transitive verb
NH	nonhuman
OBV	obviative reference
PAST	past tense
PL	plural
PROG	progressive
PROX	proximate reference
REC	recent past
SG	singular
TV	transitive verb

References

- Ambridge, B., & Lieven, E. V. M.** (2011). *Child language acquisition: Contrasting theoretical approaches*. Cambridge University Press.
- Archibald, J.** (1995). The acquisition of stress. In J. Archibald (Ed.), *Phonological acquisition and phonological theory* (pp. 81–109). Hillsdale, NJ: Erlbaum.
- Bavin, E., & Stoll, S.** (Eds.). (2013). *The acquisition of ergativity*. Amsterdam: John Benjamins.
- Berman, R.** (2014). Cross-linguistic comparisons in child language research. *Journal of Child Language*, 41 (1), 26–37.
- Bernhardt, B. H., & Stemberger, J. P.** (1998). *Handbook of phonological development: From the perspective of constraint-based nonlinear phonology*. New York: Academic Press.
- Berthiaume, S.** (2012). *A Phonological Grammar of Northern Pame*. SIL ebooks, 37. Dallas, TX: Summer Institute of Linguistics.
- Berthiaume, S.** (2019). Prosody as seen in the morphological development of Northern Pame (Xi'uiy). DFW Metroplex Linguistics Conference, November 8, 2019, Dallas, TX.
- Bickel, B., & Nichols, J.** (2007). Inflectional morphology. In T. Shopen (Ed.), *Language typology and syntactic description. Vol. 3: Grammatical categories and the lexicon*. 2nd ed., (pp. 169–240). Cambridge University Press.
- Boas, F.** (1911). Introduction. *Handbook of the American Indian Languages*, 1, 5–83.
- Bowerman, M.** (2011). Linguistic typology and first language acquisition. In J. J. Song (Ed.), *The Oxford handbook of linguistic typology*, (pp. 591–617). Oxford University Press.
- Bowerman, M., & Choi, S.** (2001). Shaping meanings for language: universal and language-specific in the acquisition of spatial semantic categories. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 475–511). Cambridge University Press.
- Brenzinger, M.** (Ed.). (2007). *Language diversity endangered*. Berlin: De Gruyter.
- Brown, P.** (1994). The INs and ONs of Tzeltal locative expressions: The semantics of static descriptions of location. *Linguistics*, 32, 743–790.
- Brown, P.** (1998). Children's first verbs in Tzeltal: Evidence for an early verb category. *Linguistics*, 36, 713–53.

- Brown, P.** (2001). Learning to talk about motion UP and DOWN in Tzeltal: Is there a language-specific bias for verb learning? In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 512–543). Cambridge University Press.
- Brown, P., Pfeiler, B., de León, L., & Pye, C.** (2013). The acquisition of agreement in four Mayan languages. In E. Bavin and S. Stoll (Eds.), *The acquisition of ergativity* (pp. 271–305). John Benjamins: Amsterdam.
- Brown, R.** (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.
- Campbell, L.** (2017). On how and why languages become endangered. *Language*, 93(4), e224–e233.
- Campbell, L., & Belew, A.** (Eds.). (2018). *Cataloguing the world's endangered languages*. London: Routledge.
- Carrillo Carreón, C.** (2007). Early acquisition of split intransitive system in Yukatek. In B. Pfeiler (Ed.), *Learning indigenous languages: Child language acquisition in Mesoamerica* (pp. 69–83). Berlin: Mouton de Gruyter.
- Clark, E. V.** (2003). *First language acquisition*. Cambridge University Press.
- Comrie, B.** (1981). *Language universals and linguistic typology*. Oxford: Blackwell.
- Cook, E.-D.** (2006). The patterns of consonantal acquisition and change in Chipewyan (Dëne Sųliné). *International Journal of American Linguistics* 72(2), 236–263.
- Crystal, D.** (2000). *Language death*. Cambridge University Press.
- Crystal, D.** (2014). Editorial. *Journal of Child Language*, 41, Supplement 1, v–vi.
- de León, L.** (2001). Why Tzotzil (Mayan) children prefer verbs: The role of linguistic and cultural factors over cognitive determinants. In M. Almgren, A. Barrena, M.-J. Ezeizabarrena, I. Idiazabal, & B. MacWhinney (Eds.), *Proceedings of the 8th conference of the international association for the study of child language* (pp. 947–969). Somerville, MA: Cascadilla Press.
- de León, L.** (2007). A preliminary view at Ch'ol (Mayan) early lexicon: The role of language and cultural context. In B. Pfeiler (Ed.), *Learning indigenous languages: Child language acquisition in Mesoamerica* (pp. 85–101). Berlin: Mouton de Gruyter.
- de López, K. M. J.** (2007). Bcuaa quiang – I stepped HEAD it! The acquisition of Zapotec bodypart locatives. In B. Pfeiler (Ed.), *Learning indigenous languages: Child language acquisition in Mesoamerica* (pp. 155–182). Berlin: Mouton de Gruyter.
- de Villiers, J., & de Villiers, P.** (1973). A cross-sectional study of the acquisition of grammatical morphemes in child speech. *Journal of Psycholinguistic Research*, 2, 267–278.
- Demuth, K.** (1996a). The prosodic structure of early words. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (171–184). Mahwah, NJ: Erlbaum.
- Demuth, K.** (1996b). Collecting spontaneous production data. In D. McDaniel, C. McKee & H. S. Cairns (Eds.), *Methods for assessing children's syntax* (pp. 3–22). MIT Press.
- Demuth, K.** (2001). Prosodic constraints on morphological development. In J. Weissenborn & B. Höhle (Eds.), *Approaches to bootstrapping: Phonological, syntactic and neurophysiological aspects of early language acquisition* (pp. 3–21). Amsterdam: John Benjamins.
- DiCanio, C., & Bennett, R.** (in press). Prosody in Mesoamerican languages. In C. Gussenhoven & A. Chen (Eds.), *The Oxford handbook of language prosody*. Oxford University Press.
- Dorian, N. C.** (2004). Minority and endangered languages. In T. K. Bhatia & W. C. Ritchie (Eds.), *The handbook of bilingualism* (pp. 437–459). Oxford: Blackwell.
- Dryer, M. S., & Haspelmath, M.** (Eds.). (2013). *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info>. Accessed on 2020-01-14.)
- Du Bois, J. W.** (1987). The discourse basis of ergativity. *Language*, 63, 805–855.
- Eberhard, D. M., Simons, G. F., & Fennig, C. D.** (Eds.). (2019). *Ethnologue: Languages of the world*. Twenty-second edition. Dallas, Texas: SIL International.
- Eisenbeiss, S.** (2006). Documenting child language. In P. K. Austin (Ed.), *Language documentation and description*, Vol. 3 (pp. 106–140). London, England: SOAS.
- England, N. C.** (1983). *A grammar of Mam, a Mayan language*. The University of Texas Press.
- England, N. C.** (2003). Mayan language revival and revitalization politics: Linguists and linguistic ideologies. *American Anthropologist*, 105, 733–743. DOI: 10.1525/aa.2003.105.4.733.
- England, N. C.** (2007). The influence of Mayan-speaking linguists on the state of Mayan linguistics. In P. K. Austin & A. Simpson (Eds.), *Endangered languages* (pp. 93–111). Berlin: Helmut Buske.

- England, N. C., & Martin, L. (2003). Issues in the comparative argument structure analysis in Mayan narratives. In J. W. Du Bois, L. E. Kumpf & W. J. Ashby (Eds.), *Preferred argument structure* (pp. 130–155). Amsterdam: John Benjamins.
- Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, 32, 429–492.
- Fee, E. J. (1995). Segments and syllables in early language acquisition. In J. Archibald (Ed.), *Phonological acquisition and phonological theory* (pp. 43–61). Hillsdale, NJ: Erlbaum.
- Fikkert, P. (1994). *On the acquisition of prosodic structure* (Doctoral dissertation). University of Leiden.
- Gehr, S. (2013). *Breath of life: Revitalizing California's native languages through archives*. San José, CA: San José State University master's thesis. Online: http://scholarworks.sjsu.edu/etd_theses/4386.
- Gómez López, P. (1998). Factores perceptuales y semánticos en la adquisición de la morfología en huichol. *Función*, 18, 175–204.
- Gómez López, P. (2007). Acquisition of referential and relational words in Huichol: From 16 to 24 months of age. In B. Pfeiler (Ed.), *Learning indigenous languages: Child language acquisition in Mesoamerica* (pp. 103–118). Berlin: Mouton de Gruyter.
- Grenoble, L. A., & Whaley, L. J. (Eds.). (1998). *Endangered languages: Language loss and community response*. Cambridge University Press.
- Grinevald, C. (2003). Classifier systems in the context of a typology of nominal classification. In Emmorey, K. (Ed.), *Perspectives on classifier constructions in sign languages*, (pp. 91–110). Mahwah, NJ: Erlbaum.
- Hale, K. (1966). Kinship reflections in syntax. *Word*, 22, 318–324.
- Hale, K. (1992). Endangered languages: On endangered languages and the safeguarding of diversity. *Language*, 68(1), 1–42.
- Hale, K. (1998). On endangered languages and the importance of linguistic diversity. In Grenoble, L. A., & Whaley, L. J. (Eds.), *Endangered languages: Language loss and community response* (pp. 192–216). Cambridge University Press.
- Hamann, C. (2002). *From syntax to discourse: Pronominal clitics, null subjects and infinitives in child language*. Dordrecht, the Netherlands: Kluwer.
- Haspelmath, M. (2010). How comparative concepts and descriptive linguistic categories are different. *Language*, 86(3), 663–687.
- Haspelmath, M. (2011). The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica*, 45(1), 31–80.
- Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. University of Chicago Press.
- Hinton, L. (2001). The use of linguistic archives in language revitalization. In L. Hinton & K. Hale (Eds.), *The green book of language revitalization* (pp. 419–423). San Diego: Academic Press.
- Hyams, N. (2011). Missing subjects in early child language. In de Villiers, J., & Roeper, T. (Eds.), *Handbook of generative approaches to language acquisition* (pp. 13–52). Dordrecht: Springer.
- Ingram, D. (1981). *Procedures for the analysis of children's language*. Baltimore, MD: University Park Press.
- Ingram, D. (1989). *First language acquisition: Method, description and explanation*. Cambridge University Press.
- Kelly, B. F., & Nordlinger, R. (2014). Fieldwork and first language acquisition. In L. Gawne & J. Vaughan (Eds.), *Selected papers from the 44th Conference of the Australian linguistic society, 2013*. Retrieved from <http://bit.ly/ALS2013Proceedings>.
- Kelly, B. F., Forshaw, W., Nordlinger, R., & Wigglesworth, G. (2015). Linguistic diversity in first language acquisition research: Moving beyond the challenges. *First language*, 35(4–5), 286–304.
- Krauss, M. E. (1992). The world's languages in crisis. *Language*, 68(1), 4–10.
- Kroeber, A. L. (1916). The speech of a Zuñi child. *The American Anthropologist*, XVIII, 529–534.
- Lahiri, A. (2001). Metrical patterns. *Language Typology and Language Universals*, 13, 47–67.
- Larsen, K. (1949). Huasteco baby talk. *El México Antiguo*, 7, 295–298.
- Lieven, E., & Stoll, S. (2013). Early communicative development in two cultures: A comparison of the communicative environments of children from two cultures. *Human Development*, 56, 178–206.
- Lust, B. (2006). *Child language: Acquisition and growth*. Cambridge University Press.
- MacWhinney, B., & Snow, C. E. (1985). The child language data exchange system. *Journal of Child Language*, 12, 271–296.
- McQuown, N. (1955). The indigenous languages of Latin America. *American Anthropologist*, 57, 501–570.

- Menn, L.** (1983). Development of articulatory, phonetic, and phonological capabilities. In Butterworth, B. (Ed.), *Language production*, vol. 2 (pp. 3–50). London: Academic Press.
- Miller, J. F.** (1981). *Assessing language production in children: Experimental procedures*. Boston: Allyn and Bacon.
- Milroy, L.** (1987). *Observing and analysing natural language: A critical account of sociolinguistic method*. New York: Basil Blackwell.
- Mithun, M.** (1989). The acquisition of polysynthesis. *Journal of Child Language*, 16, 285–312.
- Mithun, M.** (1998). The significance of diversity in language endangerment and preservation. In L. A. Grenoble & L. J. Whaley (Eds.), *Endangered languages: Language loss and community response* (pp. 163–191). Cambridge University Press.
- Mithun, M.** (2014). Morphology: What's in a word? In C. Genetti (Ed.), *How languages work: An introduction to language and linguistics* (pp. 71–99). Cambridge University Press.
- Montes Castañeda, F.** (2014). *Adquisición de la semántica de las raíces léxicas topológicas de contacto y separación: 'poner(se)' y 'quitar(se)' prendas de vestir en el totonaco de Santa Ana, Chumatlán, Veracruz*. Master's thesis. Centro de Investigaciones y Estudios Superiores en Antropología Social, México.
- Nelson, K.** (1973). Structure and strategy in learning to talk. *Monographs of the Society of Research in Child Development*, 38, no. 149.
- Nelson, K.** (2014). A matter of meaning: Reflections on forty years of JCL. *Journal of Child Language*, 41, Supplement 1: 93–104.
- Nokony, A.** (1978). Word and gesture usage by an Indian child. In A. Lock (Ed.), *Action, gesture and symbol: The emergence of language* (pp. 291–307). New York: Academic Press.
- Nolan, R.** (2020). Language barrier: For indigenous people who cross the border, a translation crisis follows. *The New Yorker*, January 6, 2020 (pp. 26–31).
- Pfeiler, B.** (2002). Noun and verb acquisition in Yucatec Maya. In M. D. Voekova & W. U. Dressler (Eds.), *Pre- and Protomorphology: Early phases of morphological development in nouns and verbs* (pp. 75–82). Lincom Studies in Theoretical Linguistics 29. München: Lincom.
- Pfeiler, B.** (2003). Early acquisition of the verbal complex in Yucatec Maya. In D. Bittner, W. U. Dressler & M. Kilani-Schoch (Eds.), *Development of verb inflections on first language acquisition. A cross-linguistic perspective* (pp. 379–399). Berlin/New York: Mouton de Gruyter.
- Pfeiler, B.** (2007). Introduction: The view from Mesoamerica. In B. Pfeiler (Ed.), *Learning indigenous languages: Child language acquisition in Mesoamerica* (pp. 1–13). Berlin: Mouton de Gruyter.
- Pfeiler, B., & Pye, C.** (2012). The acquisition of pronominal clitics in Teenek Maya (hua). Poster presentation at the 5th Conference on Generative Approaches to Language Acquisition North America (GALANA 2012), The University of Kansas, Lawrence, KS, October 12, 2012.
- Port, R., & Leary, A.** (2005). Against formal phonology. *Language*, 81, 927–964.
- Pye, C.** (1983). Mayan telegraphese: Intonational determinants of inflectional development in Quiché Mayan. *Language*, 59, 583–604.
- Pye, C.** (1990). The acquisition of ergative languages. *Linguistics*, 28, 1291–1330.
- Pye, C.** (2013). A tale of two Mam children: Contact induced language change in Mayan child language. *International Journal of American Linguistics*, 79 (4), 555–575.
- Pye, C.** (2017). *The comparative method of language acquisition research*. University of Chicago Press.
- Pye, C., Ingram, D., & List, H.** (1987). A comparison of initial consonant acquisition in English and Quiché. In K. Nelson and A. van Kleeck (Eds.), *Children's language*, vol. 6 (pp. 175–190). Hillsdale, NJ: Erlbaum.
- Pye, C., Pfeiler, B., de León, L., Brown, P., & Mateo, P.** (2007). Roots or edges?: A comparative study of Mayan children's early verb forms. In B. Pfeiler (Ed.), *Learning indigenous languages: Child language acquisition in Mesoamerica* (pp. 15–46). Berlin: Mouton de Gruyter.
- Pye, C., Pfeiler, B., & Mateo Pedro, P.** (2013). The acquisition of extended ergativity in Mam, Q'anjob'al and Yucatec. In E. Bavin & S. Stoll (Eds.), *The acquisition of ergativity*, (pp. 307–335). Amsterdam: John Benjamins.
- Pye, C., & Pfeiler, B.** (2017a). A comparative study of the acquisition of nominative and ergative agreement in European and Mayan languages. In D. Massam, L. Travis & J. Coon (Eds.), *The Oxford handbook of ergativity* (pp. 665–689). Oxford University Press.

- Pye, C., Pfeiler, B., & Mateo Pedro, P.** (2017b). The acquisition of Mayan languages. In J. Aissen, N. England & R. Zavala Maldonado (Eds.), *The Mayan languages* (pp. 19–42). London: Routledge.
- Pye, C., Pfeiler, B., Mateo Pedro, P., & Stengel, D.** (2017c). Analysis of variation in Mayan child phonologies. *Lingua*, 198, 38–52.
- Pye, C., & Pfeiler, B.** (2019). The acquisition of directionals in two Mayan languages. *Frontiers in Psychology*, 10, 2442. [doi: 10.3389/fpsyg.2019.02442].
- Pye, C., Berthiaume, S., & Pfeiler, B.** (2020). Northern Pame-Spanish language acquisition in the context of incipient language loss. *International Journal of Bilingualism*, 24(2), 246–265.
- Pye, C., Berthiaume, S., & Pfeiler, B.** (In press). The acquisition of noun inflection in Northern Pame (Xi'iuy): Comparing whole word and minimal word accounts. *Journal of Child Language*.
- Ramírez Nava, E.** (2005). *Adquisición de los prefijos de sujeto y objeto en el nahuatl de la Huasteca Veracruzana. Un estudio de caso de Hueycuatitla, Benito Juárez, Veracruz*. MA thesis. México, D.F.: Centro de Investigaciones y Estudios Superiores en Antropología Social.
- Rescorla, L. A.** (1980). Overextension in early language development. *Journal of Child Language*, 7, 321–335.
- Rice, K., & Avery, P.** (1994). Variability in a deterministic model of language acquisition: A theory of segmental elaboration. In Archibald, J. (Ed.), *Phonological acquisition and phonological theory* (pp. 33–64). Hillsdale, NJ: Erlbaum.
- Selkirk, E.** (1984). *Phonology and syntax: The relation between sound and structure*. MIT Press.
- Siewierska, A.** (2004). *Person*. Cambridge University Press.
- Slobin, D. I.** (Ed.). (1967). *A field manual for cross-cultural study of the acquisition of communicative competence*. Berkeley, CA: Language-Behavior Research Laboratory, University of California.
- Slobin, D. I.** (1985). Introduction: Why study acquisition crosslinguistically? In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition. Vol. 1: The data* (pp. 3–24). Hillsdale, NJ: Erlbaum.
- Stassen, L.** (2011). The problem of cross-linguistic identification. In J. J. Song (Ed.), *The Oxford handbook of linguistic typology* (pp. 90–99). Oxford University Press.
- Stemberger, J. P., & Chávez-Peón, M. E.** (2014). Overgeneralization in the processing of complex forms in Valley Zapotec child language. *Semantic considerations of lexical processing. Thematic issue of the mental lexicon*, 9(1), 107–130.
- Stoll, S.** (2016). Studying language acquisition in different linguistic and cultural settings. In N. Bonvillian (Ed.), *The Routledge handbook of linguistic anthropology* (pp. 140–158). New York: Routledge.
- Stoll, S., & Lieven, E.** (2014). Studying language acquisition cross-linguistically. In H. Winkler & P. Pradakannaya (Eds.), *South and Southeast Asian psycholinguistics* (pp. 19–35). Cambridge University Press. doi:10.1017/cbo9781139084642.004.
- Straight, S. H.** (1976). *The acquisition of Maya phonology: Variation in Yucatec child language*. New York: Garland.
- Stross, B.** (1969). Language acquisition by Tenejapa Tzeltal children. *Working paper No. 20*. Language-Behavior Research Laboratory, University of California, Berkeley.
- Thomason, S.** (2015). *Endangered languages: An introduction*. Cambridge University Press.
- Thomason, S. G., & Kaufman, T.** (1988). *Language contact, creolization, and genetic linguistics*. University of California Press.
- van Gijn, R., & Zúñiga, F.** (2014). Word and the Americanist perspective. *Morphology* 24 (Special Issue on Word in the Languages of the Americas), 135–160.
- Wallace, A. R.** (1863). On the physical geography of the Malay archipelago. *Journal of the royal geographical society*, 33, 217–234.
- Wright, S. H.** (2003). Professor brings Aboriginal languages to life. In MIT News, 22 October 2003.