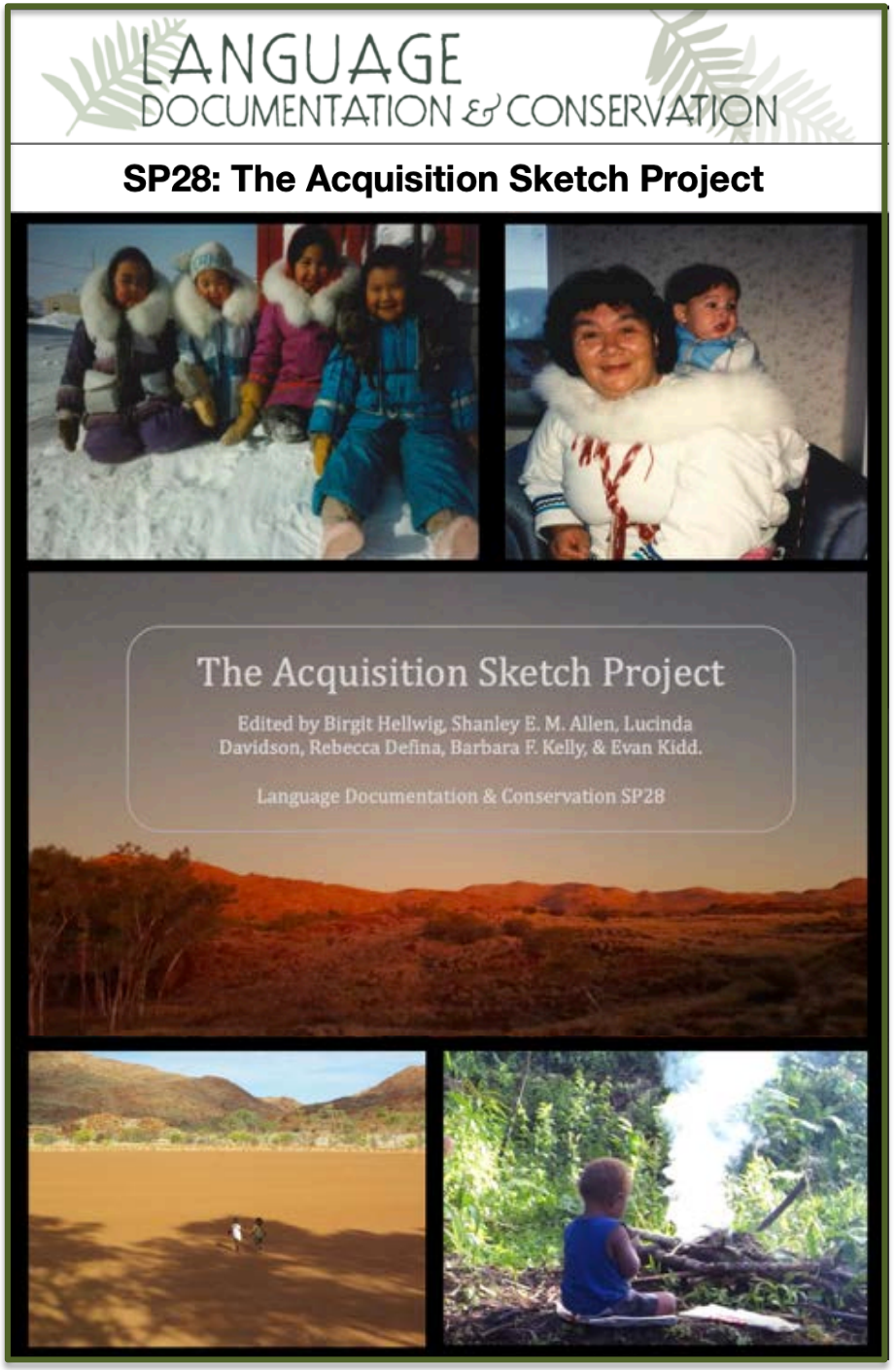
THE ACQUISITION SKETCH PROJECT:

INCREASING DIVERSITY IN LANGUAGE ACQUISITION RESEARCH

Bimonthly Workshop

May 17, 2023



TEAM AND FUNDERS

Shanley Allen

Lucy Davidson

Rebecca Defina

Birgit Hellwig

Barbara Kelly

Evan Kidd

DAAD

Deutscher Akademischer Austauschdienst
German Academic Exchange Service

UNIVERSITIES AUSTRALIA

MOTIVATION

GOAL OF LANGUAGE ACQUISITION RESEARCH

To understand how humans acquire the capacity to perceive, comprehend, and produce language

Biological factors …

- *Innate linguistic and/or cognitive abilities*
- *Mental architecture, mechanisms of storage and access*

Environmental factors …

- *Input, interaction*
- *Culture, family, SES*

Linguistic factors …

- *Language structure – phonology, morphology, syntax, semantics, etc.*
- *Frequency, integrating multiple cues*

Need data from many languages to understand what is really universal (i.e. true for all humans) and what is language-specific

IMPORTANCE OF MANY LANGUAGES

"[T]o show how the child's mind can learn …, with approximately equal ease, any one of [a] vast range of alternative systems … calls for a **diversified and strategic harnessing of linguistic diversity** as the independent variable in studying language acquisition …: Can different systems be acquired by the same learning strategies, are learning rates really equivalent, and are some types of structure in fact easier to use?"

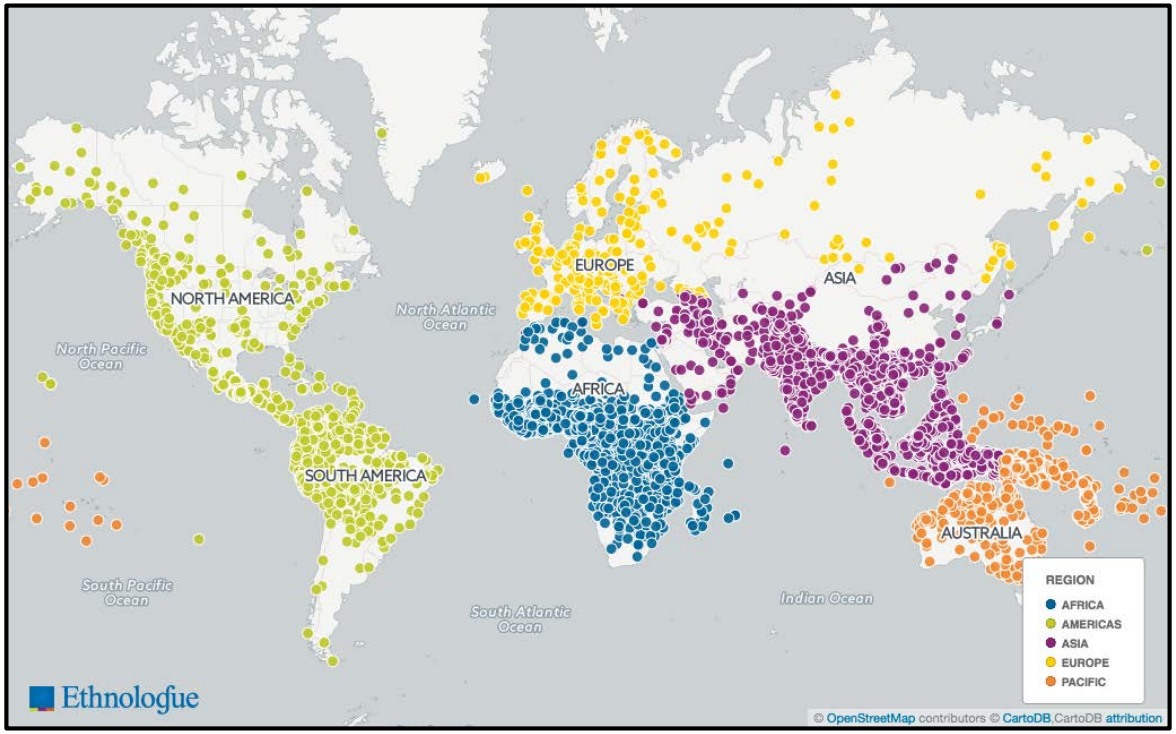
Evans & Levinson 2009:447

VAST LINGUISTIC DIVERSITY

7000+ languages in the world, 400+ language families

Represents vast diversity in linguistic phenomena and structures

50% of these languages predicted to be lost by end of century



EXISTING RICH TRADITION

Rich tradition of crosslinguistic work including:

- Comparable analysis across languages (Slobin 1985-1997)
- Comparative analysis across languages
  (e.g. Berman & Slobin 1994, Dressler & Karpf 1995, MacWhinney & Bates 1989)
- Data archiving, sharing, and analysis programs (MacWhinney 2000)
- Individual corpora in several languages
  (e.g. K'iche (Pye), Sesotho (Demuth), Zulu (Suzman), Warlpiri (Bavin), Inuktitut (Allen), Mohawk (Mithun), Navajo (Courtney)

Clear impact of this work on theory and understanding of acquisition

BUT STILL MUCH TO BE DONE

<2% of languages covered in top 4 acquisition journals (103/7000)

Research highly skewed to English and other Indo-European languages

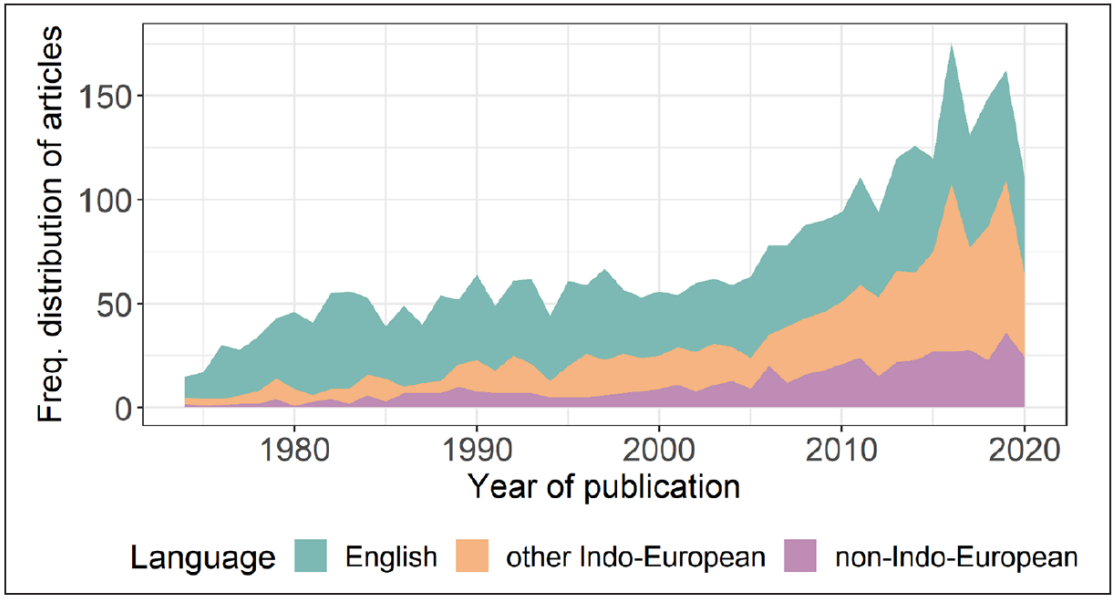


Figure 1. Frequency Distribution of Articles on English, other Indo-European Languages and Non-Indo-European Languages Published in Child Language Journals Between 1974 and 2020.

English: 54%, Other IE: 30%, Non-IE: 16%

Kidd & Garcia 2022

BUT STILL MUCH TO BE DONE

Indo-European and non-IE research highly skewed to a few languages

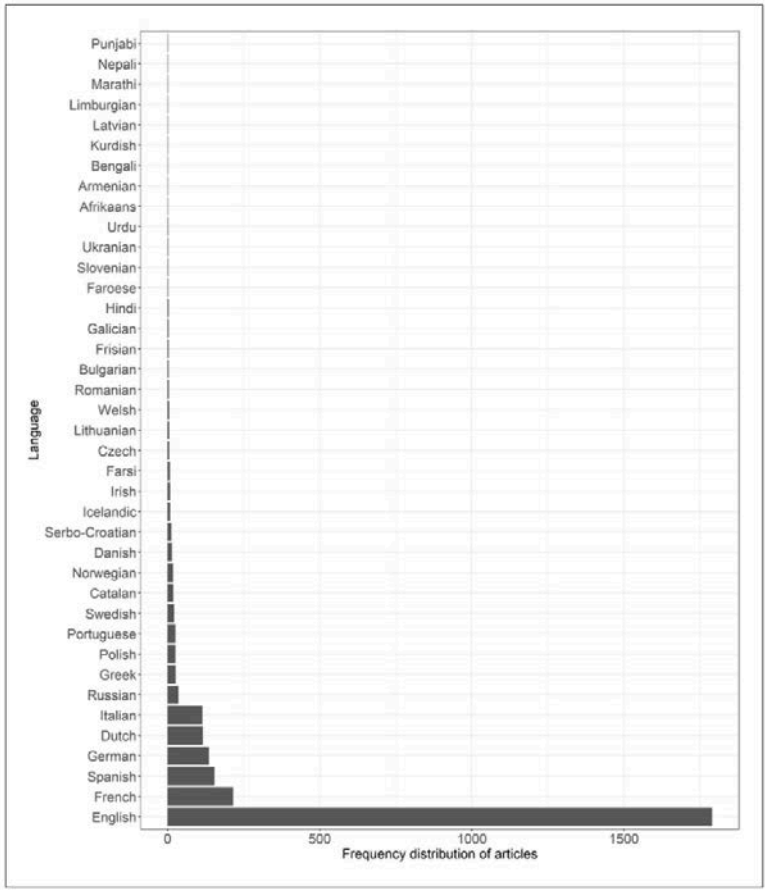


Figure 2. Frequency Distribution of Articles on Indo-European Languages Published in Child Language Journals Between 1974 and 2020.

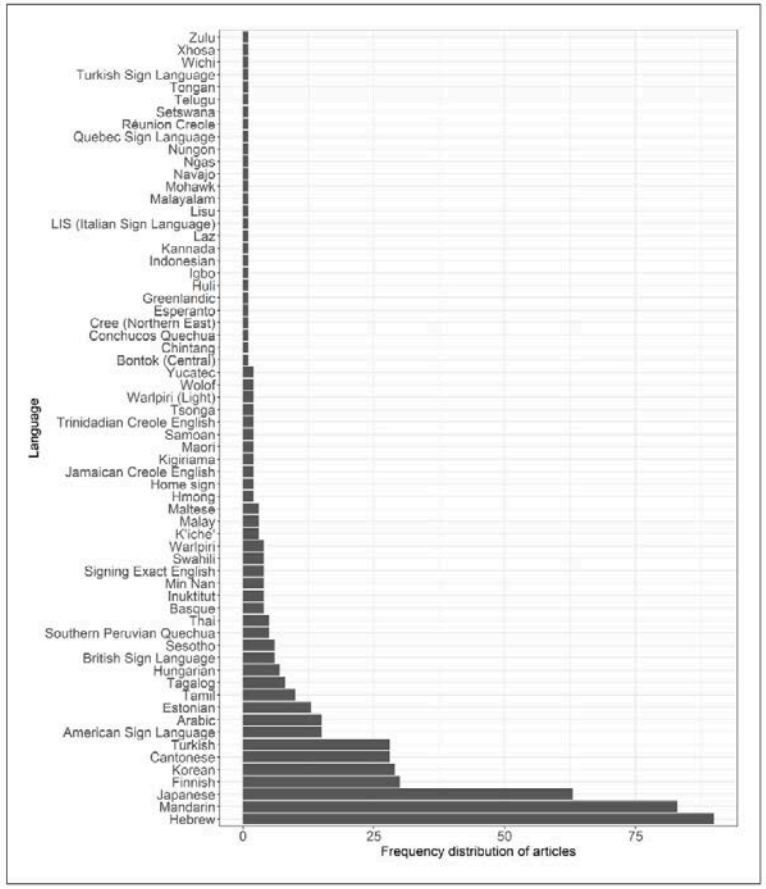


Figure 3. Frequency Distribution of Articles on Non-Indo-European Languages Published in Child Language Journals Between 1974 and 2020.

Kidd & Garcia 2022

KEY QUESTION

How can we tractably study the acquisition of minority and endangered languages in a way that responds to both rapid endangerment and the need to widen the evidential base of the field?

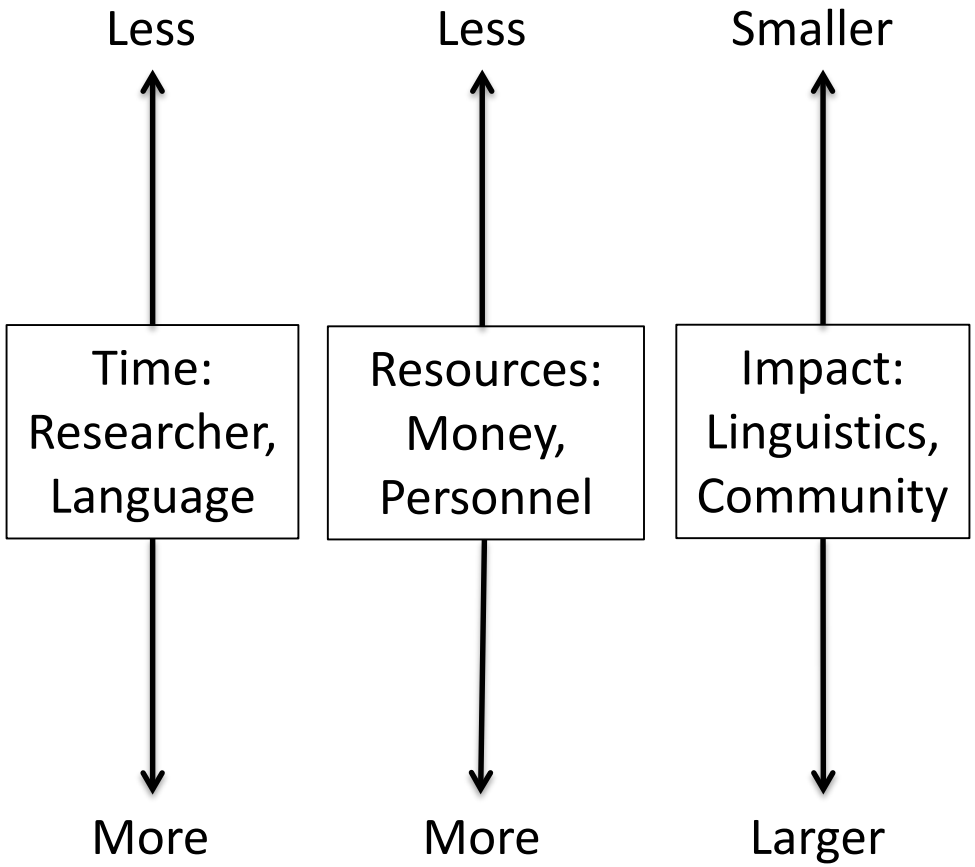
Defina et al. 2023a, p. 6

WAYS TO ADDRESS THE CHALLENGE

Word list for one child

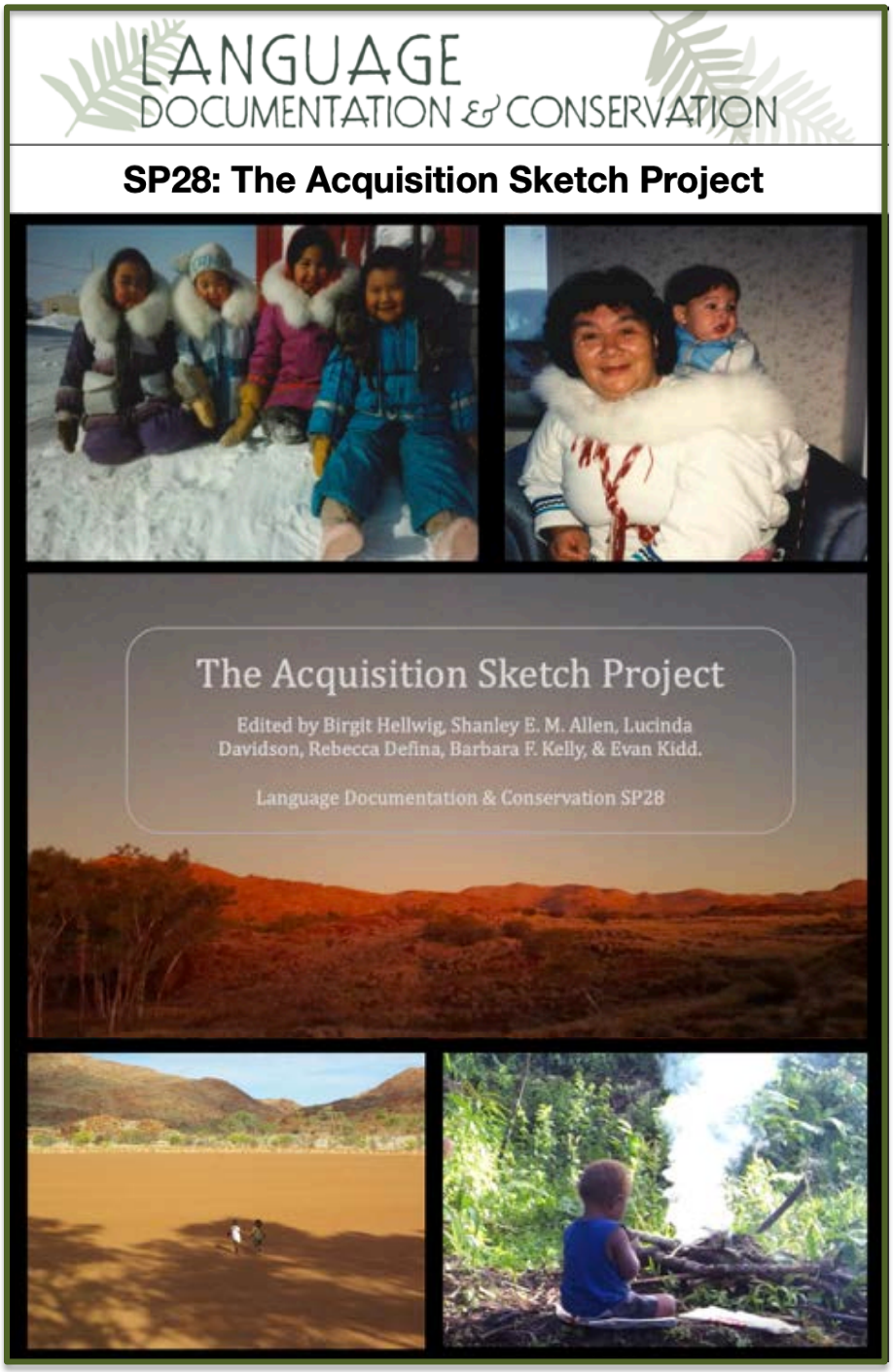
Solutions that are both useful and doable!

Full longitudinal study for 20 children



→ Need different approaches for different situations

ACQUISITION SKETCH PROJECT



https://nflrc.hawaii.edu/ldc/sp28-the-acquisition-sketch-project/

GUIDING FOCUS FOR PROJECT

Facilitate collection and analysis of acquisition corpora that are:

- Small-scale and manageable
- Comparable to facilitate crosslinguistic comparison
- Achievable with realistic resources
- Useful for testing and developing theories
- Useful for documentation, maintenance, and revitalization
- Well-annotated and sustainably archived
- Incorporating ethnographic and metalanguage data

Defina et al. 2023

KEY FEATURES OF ACQUISITION SKETCH

Observable spontaneous language use in dynamic everyday contexts

Main age range for development (ages 2, 2½, 3, 3½, 4)

Representative-ish – two children at each age

Five hours of data – one hour for each age (30 minutes per child)

Main areas of language
– phonology, lexicon, morphology, syntax, gesture

Child language and child-directed language

Defina et al. 2023

LIKELY SKETCH WRITERS

Researchers with:

- Experience and knowledge of the language
- Some resources and time
- Community with fluent L1 learners
- Not necessary to have acquisition background

Ideal for:

- Language documenters
- Community members
- PhD students

Goal: Reduce barriers via concrete guidelines

SKETCH ACQUISITION MANUAL (SAM)

SP28 (2023), pp. 5–38
http://nflrc.hawaii.edu/ldc
https://hdl.handle.net/10125/74719

The Acquisition Sketch Project

Sketch Acquisition Manual (SAM), Part I: The sketch corpus

Sketch Acquisition Manual (SAM), Part II: The acquisition sketch

Rebecca Defina,[1,2] Shanley E. M. Allen,[3] Lucinda Davidson,[1,2] Birgit Hellwig,[4] Barbara F. Kelly,[1] Evan Kidd [2,5,6]

Defina et al. 2023a, b

PART 1: THE SKETCH CORPUS

CONSTRUCTING THE CORPUS

Structure of the Corpus

- ❑ Age/number of children
- ❑ Amount of data
- ❑ Participants and content
- ❑ Rationale for setup

Practical Considerations

- ❑ Identifying children and contexts
- ❑ Recording setup
- ❑ Archiving and metadata
- ❑ Ethical considerations

Defina et al. 2023a

PART 1: THE SKETCH CORPUS
PROCESSING THE DATA

Preparing Files for Transcription

- ❑ Transcribers
- ❑ Segmentation
- ❑ Tiers

Transcription and Translation

- ❑ Deciphering utterances
- ❑ Transcription and the adult interpretation
- ❑ Transcription as data collection
- ❑ Beyond transcription

Defina et al. 2023a

GUIDELINES AND SUGGESTIONS:
DATA RECORDING OPTIONS

Three possible recording options are outlined in Table 3. Ideally, try to record each child over the course of an entire day. Such a setup will give you a good record of the learning environment of the children: their daily routines, the activities they engage in, and the type and amount of their interaction with adults and other children. Another option is to record for multiple hours within the same week, so that you capture a fair amount of the language that children encounter and produce at that point in time, with a chance of picking up low-frequency phenomena. Alternatively, record a minimum of 60 minutes.

Table 3. Amounts of data to be recorded and processed.

	Recording (per child and age)	Processing (per child and age)
Ideal	Day-long recording	30 minutes
Alternative	3-5 (or more) hours within same week	30 minutes
Minimum	60 minutes	30 minutes

Defina et al. 2023a, p. 11

QUESTIONS TO SPARK IDEAS:
SELECTING PARTICIPANTS

First, we recommend discussing the following three points with as many community members as appropriate:

- Which families have children of the right ages? Does the community consider them suitable families for the project (e.g. in terms of social standing, language background and use, etc.)? Do the families themselves have the time and inclination to participate?
- How do people assess and talk about the linguistic and non-linguistic development of children? Which linguistic structures and/or non-linguistic skills are children expected to master at which ages? Which specific children master the identified structures and skills?
- What do children of different ages typically do during various parts of their day? In which contexts do they interact with others, in which not? With whom do they interact?

Defina et al. 2023a, p. 14

EXAMPLES FROM EXPERIENCE:
EQUIPMENT / TRANSCRIBING OPTIONS

Figure 1. Example of recording setup (photos: Lucinda Davidson), showing the location of the transmitter and hidden lapel microphone in the child's backpack, and the receiver plugged into the camera. Children are free to run about while wearing the backpacks.

Figure 3b. Tier setup in the Qaqet corpus: Toolbox.

LongZDL20160112_2.txt

\ref reference	LongZDL20160112_2 0638
\ELANBegin ELAN Begin	897.471
\ELANEnd ELAN End	898.139
\ELANParticipant Participant	ZDL
\addr addressee	APA
\trs transcript (actual utterance)	**pupuka**
\trs-i transcript (interpretation)	**apupuqa**
\ft free translation (English)	grandpa
\tp free translation (Tok Pisin)	*pupuman*
\tx text	**pupuka**
\mb morpheme breaks	**pupu -ka**
\ge gloss (English)	granny -NC.SG.M
\ps part of speech	N -sfx
\lg language	TP -Q
\idx index	01344 -00732
\nt notes	
\sound sound	LongZDL20160112_2.wav 897.471 898.139
\dt date	20/May/2019

Figure 3c. Tier setup in the Qaqet corpus: CHAT.

```
*ZDL:    pupuka
%int:    apupuqa
%eng:    grandpa
%tkp:    pupuman
%mor:    n|pupu=grandparent&TP+nc|ka=sg&m&Q
%pos:    n+sfx
%add:    APA
%not:
%snd:    LongZDL20160112_2.wav 897.471 898.139
```

Defina et al. 2023a, p. 17, p. 28

PART 2: THE ACQUISITION SKETCH

CONTEXTUALIZING THE SKETCH

Background

- ❑ Language typology
- ❑ Data
- ❑ Learning environment and ethnotheories
- ❑ Multilingualism

Outlook

- ❑ Community report

Defina et al. 2023

PART 2: THE ACQUISITION SKETCH

WRITING THE SKETCH

Child-Directed Language

- ❑ Scaffolding techniques
- ❑ Responding to non-target-like utterances
- ❑ Phonology and prosody
- ❑ Lexicon and semantics
- ❑ Morphosyntax
- ❑ Gestures

Child Language

- ❑ Phonology and prosody
- ❑ Lexicon and semantics
- ❑ Morphology
- ❑ Syntax
- ❑ Gesture

Defina et al. 2023

GUIDELINES AND SUGGESTIONS:
CHILD MORPHOLOGY

6.3 Morphology

Languages differ widely in their morphological complexity (e.g. isolating vs. polysynthetic, agglutinative vs. fusional), and each type of language presents different challenges for learning morphology. Thus, you should begin by thinking carefully about the morphology of the language and what challenges it is likely to present. The morphological system is also very likely to affect patterns of acquisition. For example, children are faster and show fewer non-target-like forms learning inflectional systems that are systematic (e.g. Turkish, Inuktitut) than systems that are not systematic (i.e. with many empty cells like English, or with lots of syncretism like German), regardless of the relative complexity of the system. In addition, very young children tend to reduce morphologically complex target words (see Section 6.1 on word structure).

Regardless of the language typology, important morphological developments will take place during the ages covered by the sketch corpus. For example, the first instances of inflectional morphology will likely start to appear around age 2;0 and become more frequent and regularized over the following age points. The main challenges will be to identify the morphemes of each type at each stage, and then to hypothesize about possible developments between stages. Depending on what is relevant for the language, we suggest focusing on:

- Nominal morphology: case, number, gender, noun class, etc.
- Verbal morphology: agreement, tense, aspect, evidentiality, etc.
- Grammatical morphology: passive, antipassive, causative, negation, etc.

Defina et al. 2023b, p. 69

KEY CONCEPTS FROM LITERATURE:
TYPICAL FEATURES OF CHILD-DIRECTED LANGUAGE

Table 2. Overview: Typical features of child-directed language.

short (but complete) utterances
few hesitations and errors
exaggerated pitch contours, larger and more clearly articulated signs
high F0
long duration and pauses
restricted vocabulary with reference to the present time and location
nursery vocabulary
many questions, imperatives and prompts
many full and partial repetitions, variation sets

Defina et al. 2023b, p. 50

QUESTIONS TO SPARK IDEAS:
ROUTINES IN CHILD-CAREGIVER INTERACTION

Routines

- Are there any salient routines when talking to children? For example, anthropological research often reports on practices that can be considered functional equivalents of child-directed language, e.g. the *a:la:ma* (also known as *elema*) 'say like that' routine of the Kaluli where adults model the child's utterance and prompt the child to repeat it (Ochs & Schieffelin 1996: 86-87; see also example 1). If you observe any such routines, do children engage in them as well and use the same kinds of routines in their interaction with others?
- Do adults prompt children as a means of teaching appropriate pragmatic behavior such as by modeling "say thank you" or by framing utterances telling a child what to say, such as "say, 'I don't like it'".
- Are there any other interactional routines? For example, what do people do to soothe or comfort children? To distract them? To entertain them? To tease them? To prevent them from getting hurt? Are there differences in routines with strangers vs. with familiar people?
- Are there any songs/chants/lullabies that are commonly used? Any games?

Defina et al. 2023b, p. 51

SUMMARY BOXES:
KEY AREAS OF FOCUS IN PHONOLOGY

Box 8. Key areas of focus in child language: Phonology and prosody.

Core

(i) What is the distribution of phonemes, tonemes and syllable structures across development?

(ii) Do you observe any (target- or non-target-like) phonological processes? Which ones and at which ages?

(iii) Do you observe any patterns in the truncation of words? Which ones and at which ages?

Extension

(iv) Expand on (iii): investigate children's use of stress.

(v) Do you observe any salient prosodic patterns in child language? Which ones?

Defina et al. 2023b, p. 67

EXAMPLES FROM ONGOING SKETCHES:
DATA SET FOR PITJANTJATJARA

Table 1. Data set for the Pitjantjatjara sketch. The table lists the age bracket, the focus child ID reference, the number of utterances produced by the focus child, the number of other children present at the recording, the number of utterances produced by those other children, the number of adults present, and the number of utterances produced by those adults.

Age	Focus child	Focus child utterances	No. of other children	Other child utterances	No. of adults	Adult utterances
2;0	ANT	176	3	409	3	283
	ANN	177	5	399	5	281
2;6	ANT	230	4	355	3	229
	FRE	194	1	139	1	217
3;0	ANT	219	2	283	3	329
	REN	298	1	5	4	567
3;6	ISY	202	5	395	2	162
	REN	171	3	363	2	127
4;0	ISY	212	3	297	8	470
	FRE	227	3	297	8	470

Defina et al. 2023b, p. 45

EXAMPLES FROM ONGOING SKETCHES:
CDL INTONATION IN QAQET

For example, we conducted such spot-checks on polar questions in the Qaqet sketch corpus. We annotated the corpus for speech acts (on a separate tier labeled "speech acts"), extracted intonation units that contained polar questions, and inspected their pitch contours in Praat (Boersma & Weenink 2022). Regardless of addressee, all Qaqet polar questions are characterized by a distinctive rise-fall contour, but those units addressed to children (up to age 3;6) exhibit exaggerated pitch movements (illustrated in Figure 1a) compared to polar questions addressed to older children and adults (illustrated in Figure 1b).

Figure 1a. Polar question addressed by AMT to YDS (2;0): *kua nyinarli?* 'do you hear?' (LongYDS20150506_1 1167.752 1168.561).

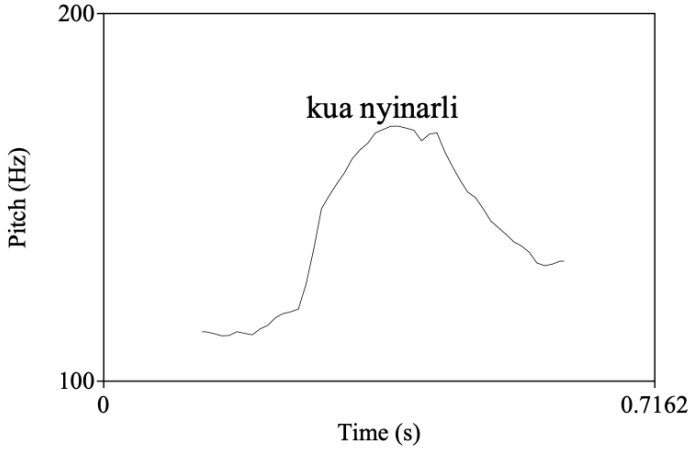
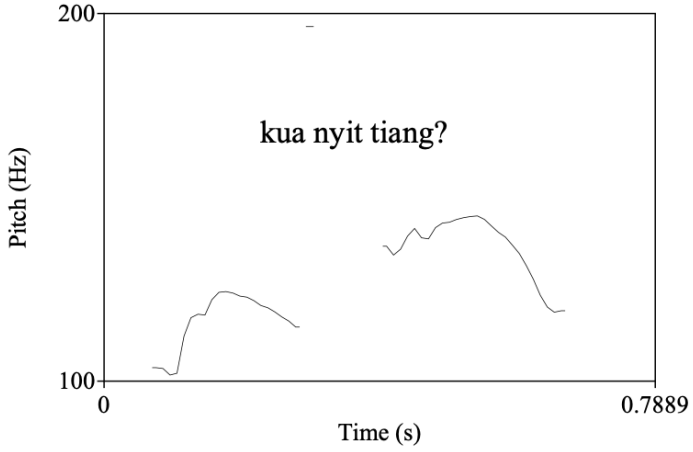


Figure 1b. Polar question addressed by AMT to ZVI (adult): *kua nyit tiang?* 'do you get others?' (LongYDS20151204_1 1001.330 1002.025).



Defina et al. 2023b, p. 56

EXAMPLES FROM ONGOING SKETCHES:
WORD TRUNCATION IN INUKTITUT

In the Inuktitut example in (5), the child is kissing a character on the TV. The sister comments on this using various affixes appended to the verb root, while the child keeps responding with simply the verb root.

(5) Inuktitut: Child (Jini 1;4) and her sister

Sister: *maa-li-ruk.*
kiss-POL-IMP.2SG.SBJ>3SG.OBJ
'Kiss her.'

Child: *maa.*
kiss
'Kiss.'

Sister: *maa-pait.*
kiss-PAR.2SG.SBJ>3SG.OBJ
'You kissed her.'

Child: *maa.*
kiss
'Kiss.'

Sister: *maa-tau-laur-langa.*
kiss-PASS-POL-IMP.1SG.SBJ
'Kiss me.' [lit: let me be kissed]

Defina et al. 2023b, p. 65

EXAMPLES FROM ONGOING SKETCHES:
DEICTIC GESTURES IN MURRINHPATHA

Figure 2a. Adult pointing gesture

Figure 2b. Child pointing gesture

Defina et al. 2023b, p. 65

THREE SKETCHES SO FAR

Broadening language documentation with child language data: First-hand experience from Tolitoli

Christoph Bracks,[1,2] Maria Bardají i Farré [2]

Exploring case marking in German first language acquisition using the acquisition sketch approach

Gianna Urbanczik

An acquisition sketch of Inuktitut

Hannah Lee, Shanley E. M. Allen

DOES MLU INCREASE WITH AGE?

Table 35. Average MLUm in children's speech.

	1;4	1;10	2;4	2;10	3;4
CHI A	1.19	1.48	2.59	3.07	3.53
CHI B	1.18	1.68	2.11	2.89	2.96
Average	1.18	1.58	2.35	2.98	3.25

MLU in morphemes increases with age

Some differences between children at a given age

- Literature: Consistent with pattern from other languages
- Larger study: Same pattern but more variation between children

Lee & Allen 2023, p. 189;
Allen & Dench 2015

HOW DOES POLYSYNTHESIS DEVELOP?

Table 42. Word length in morphemes by age in children's speech.

Group	1	2	3	4	5	6	7
1;4	0.84	0.15	0.01				
1;10	0.68	0.23	0.07	0.01	0.01		
2;4	0.40	0.35	0.15	0.06	0.04		
2;10	0.36	0.29	0.17	0.11	0.04	0.01	0.01
3;4	0.34	0.31	0.20	0.09	0.05	0.01	0.01

The number of words at each age is as follows: 1;4=73, 1;10=198, 2;4=327, 2;10=570, 3;4=441.

1-morpheme words most frequent at every age
3-morpheme words common by 2;4, 7-morpheme words by 2;10

- Literature: Consistent with West Greenlandic data
- Larger study: Similar results

Lee & Allen 2023, p. 199;
Allen & Dench 2015

DO PASSIVES EMERGE LATE?

Number of short and full passives at each age

	1;4	1;10	2;4	2;10	3;4
Short Passive	0	0	1	4	5
Full Passive	0	0	0	1	0

11 passives in 1166 utterances = 0.1%
Infrequent but present, including full passive
Present from 2;4 – much earlier than age 4-9 as predicted by UG

- Literature: Contradicts maturation, supports frequency
- Larger study: Same pattern but much more detailed analysis (productivity, verb types, frequency, interaction with structures)

Lee & Allen 2023, p. 197;
Allen & Crago 1996

DISCUSSION

SUMMARY

Acquisition Sketch Project provides:

- Format and guidelines for collecting and archiving data
- Format and guidelines for writing an acquisition sketch
- Many suggestions, tips, and examples from experience
- Clear models to follow from existing sketches
- Publishing outlet – Special Issue 28 in *LDC*
- Possibility for future comparative work across sketches
- Support from experienced child language researchers
- Community of fellow sketch writers

CAVEATS

Data are limited

- More suited for description than explanation, for breadth than depth
- Can't confirm absence of phenomenon, show productivity, do inferential statistics to confirm generalizability

Easier to see patterns if you know what to look for

- Think carefully about features of your language and how they may address current debates or gaps in the field
- Ask community members about typical strengths / difficulties

Children are variable

- Some results may depend heavily on which children you select

Work not trivial

- Much less work than larger project but still will take ± 1 year

TAKE-HOME MESSAGE

Five hours of data in an acquisition sketch can contribute a lot!

- Useful descriptive findings
- Generalizations about higher-frequency phenomena
- Promising hypotheses for further research
- Support for community materials
- Broader understanding of the variation space
- Increased diversity and corrected bias in acquisition research

Good luck in developing your own acquisition sketch!

THANK YOU!
QUESTIONS?