Australian Journal of Linguistics

ISSN: 0726-8602 (Print) 1469-2996 (Online) Journal homepage: www.tandfonline.com/journals/cajl20

A comparative study of child-directed language across five cultures based on data from the *Acquisition Sketch Project*

Evan Kidd, Birgit Hellwig, Rowena Garcia, Rebecca Defina, Lucinda Davidson & Shanley Allen

To cite this article: Evan Kidd, Birgit Hellwig, Rowena Garcia, Rebecca Defina, Lucinda Davidson & Shanley Allen (2025) A comparative study of child-directed language across five cultures based on data from the *Acquisition Sketch Project*, Australian Journal of Linguistics, 45:3, 391-415, DOI: 10.1080/07268602.2025.2514825

To link to this article: https://doi.org/10.1080/07268602.2025.2514825

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 22 Jul 2025.

Submit your article to this journal

Article views: 1110

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=cajl20

AUSTRALIAN JOURNAL OF LINGUISTICS
2025, VOL. 45, NO. 3, 391–415
https://doi.org/10.1080/07268602.2025.2514825

OPEN ACCESS Check for updates

A comparative study of child-directed language across five cultures based on data from the *Acquisition Sketch Project*

Evan Kidd, Birgit Hellwig, Rowena Garcia, Rebecca Defina, Lucinda Davidson and Shanley Allen

Australian National University, University of Cologne, Leibniz-Centre General Linguistics & University of the Philippines Manila, University of Melbourne, University of Melbourne and University of Kaiserslautern-Landau

ABSTRACT

Throughout the history of child language acquisition research, the study of child-directed language (CDL) has attracted significant attention. In particular, there has been considerable debate regarding the characteristic features of CDL and their universality/variability across the world's languages. Yet, although data from many languages have been analyzed, the totality of the crosslinguistic coverage is still poor. In this paper, we report on an analysis of CDL across five diverse languages and cultures: Murrinhpatha (Southern Daly, non-Pama-Nyungan), Pitjantjatjara (Pama-Nyungan), Qaqet (Baining), Tagalog (Western Austronesian), and Inuktitut (Inuit-Yupik-Unangan). Using data collected for the *Acquisition Sketch Project*, an initiative in which Barb was a core member, we find both striking commonalities and clear differences in CDL across our target languages. The findings are consistent with the argument that CDL emerges as a set of culturally mediated behavioural practices, with some features being more commonly observed than others. The findings underline the value of the *Acquisition Sketch* approach in widening the evidence base of the field of child language acquisition, one of Barb's major contributions to the field.

ARTICLE HISTORY
Received 19 March 2025
Accepted 16 April 2025

KEYWORDS
Child-directed language; *Acquisition Sketch Project*; Australian languages; Papuan languages; Austronesian languages; Inuit-Yupik-Unangan languages

1. Introduction

One of Barb's core beliefs about language acquisition was that children acquire language within rich social contexts. True to the functionalist tradition of her training in California, her work on acquisition investigated how social and cultural context framed communicative exchange, from which language naturally emerged across development (e.g. Davidson & Kelly, 2021; Kelly, 2006). She was a keen observer of interaction, seeing rich meaning in exchanges that many would pass over as unremarkable.

One truism of acquisition is that children are socialized into language and culture (Ochs & Schieffelin, 1984). In broad terms, this means that children are treated differently

CONTACT Evan Kidd evan.kidd@anu.edu.au

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

to competent speakers of the linguistic community because they are still learning its linguistic and cultural conventions. How this notion of difference plays out in universal and culturally specific terms is less clear. Cultural conventions and beliefs relating to child rearing and socialization, the focus of anthropology, show great variation. However, within linguistics and the broader study of child language there has been significant effort spent determining: (i) whether there are universal properties in how children are spoken to; and (ii) what functions modifications in language directed to children might have. This is the study of *child-directed language* (henceforth CDL) in a broad sense, sometimes conceived of as a special register, sometimes more generally couched in terms of input. It is directly related to and has many overlapping features with *infant-directed language*, which is typically used to refer to language directed towards very young infants. Starting with the seminal volumes by Snow and Ferguson (1977) and Gallaway and Richards (1994), the field has been identifying and discussing CDL features that are widespread across many better documented languages. In our contribution, we have selected a subset of these and explore their presence across five very different cultural contexts. In doing so we provide the first comparative study emerging from the *Acquisition Sketch Project* (Hellwig et al., 2023a), a major initiative aimed at broadening language coverage in child language research, to which Barb was a major contributor.

2. Child-directed language

CDL can be best described as a cluster of linguistic behaviours across different areas of language. Techniques to scaffold children into more target-like communication patterns include a greater frequency of speech acts like imperatives and questions, as well as prompts to mimic certain speech patterns. Typical prosodic modifications include higher and greater variation in pitch, lengthening of vowels, and greater number of pauses. Lexical modifications include phonological changes and simplification (e.g. *widdle bunny wabbit*), diminutives (e.g. *blankie* 'blanket', *doggie* 'dog'), reduplication (e.g. *dum-dum* 'dummy/pacifier'), and other *nursery vocabulary*, whether it be widespread throughout a community or particular to families. Finally, syntactic modifications include short and grammatically correct utterances with few hesitations, and the greater use of repetition and *variation sets* (a term coined by Küntay & Slobin, 1996) – partial or varied repetition of utterances that may provide opportunities for children to learn linguistic elements such as morphs or words and their distributional behaviour.

The functions of CDL are not well understood but are likely multifaceted. For instance, apart from the suggestion that CDL, and more specifically infant-directed language, may facilitate language acquisition via the provision of learnable input, it has also been proposed that CDL may better pique and hold the child's attention (see Frank et al., 2020) and may serve to foster affective bonds between the child and caregivers (for reviews see Golinkoff et al., 2015; Soderstrom, 2007).

There are important caveats related to these features and functions of CDL. Most notably, although crosslinguistic studies exist (e.g. Fernald et al., 1989), the behaviours have been observed in only a fraction of the world's languages (see next section), limiting our understanding of the variation space. For example, Bernstein Ratner and Pye (1984) reported that caregivers of children acquiring Quiché Mayan do not raise their pitch when speaking to two-year-olds, but instead talk in a lower "crooning" tone (see also Ingram, 1995). More generally, the conditions under which children acquire language cross-culturally often differ from the typically studied Western dyadic model. For example, children in non-Western cultures are regularly raised in polyadic contexts where multi-party conversation is the norm, resulting in less language being specifically directed towards young children (e.g. Casillas et al., 2020). Additionally, many cultures have been reported to not consider children conversational partners until they have already begun to talk (e.g. Kaluli: Schieffelin, 1985; Quiché Mayan: Pye, 1986; Samoan: Ochs, 1985; Warlpiri: Bavin, 1992; for review see Lieven, 1994). Thus, although children are socialized into language and culture, the conditions under which they learn language vary across many dimensions. In this respect, we conceive of CDL in terms of family resemblance; that is, as a cluster of behaviours that individuals and cultural groups variably draw upon across sociolinguistic and cultural contexts. The question then becomes one of identifying which behaviours feature more or less commonly across languages and cultures. In the current paper we do this using acquisition sketch data, to which we now turn.

3. The *Acquisition Sketch* approach

On present estimates, there are around 7,000 or so languages currently spoken on the planet (Eberhard et al., 2024). Natural language exhibits remarkable diversity (Evans & Levinson, 2009), so an adequate theory of language and how it is acquired must draw from a representative sample. However, we currently lack good coverage when it comes to child language data. In a study that analyzed all papers published in the four major child language journals (*Journal of Child Language*, *First Language*, *Language Acquisition*, *Language Learning and Development*) between 1974 and 2020, Kidd and Garcia (2022) found papers on only 103 languages, representing a mere 1.5% of the world's languages (see also Kelly & Nordlinger, 2014). Worse still, the majority of papers were published on English and a handful of languages spoken mostly in Western Europe, with most understudied languages containing only one paper on their acquisition. More data from diverse languages are needed, a situation that has some urgency given the rapid rate of language loss (Bromham et al., 2022; Passmore et al., 2025).

In response to this need, Defina et al. (2023a, 2023b) developed the *Acquisition Sketch* approach. Combining ideas from the fields of language documentation, language socialization, and child language acquisition, the initiative aims to broaden child language data coverage in a tractable and accessible way. In the approach, researchers write an acquisition sketch of a language based on a minimum of five hours of naturalistic data recorded with children aged around 2;0–4;0 years, a period of acquisition where development proceeds at a rapid rate. On the basis of the data, researchers analyze aspects of CDL and major developments in the children's language, focusing on broad areas such as phonological, lexical, and morphosyntactic development. Rather than being a comprehensive account of acquisition in the target language, the sketch is a "first look", serving a broad-spectrum overview that in the best cases will identify important phenomena to follow up in bigger studies. It also serves an important applied purpose: having an idea of how children acquire a language can have valuable educational and clinical functions, which is especially important when communities do not have any resources or assessment tools in the local language.

In this paper we take data from five sketch corpora to provide a comparative analysis of CDL, targeting the extent to which selected features are present or absent. The languages we examine are diverse and understudied, and are acquired within very different cultural contexts. The languages include two from Australia: Murrinhpatha (Southern Daly, non-Pama-Nyungan) and Pitjantjatjara (Pama-Nyungan); Qaqet (Baining), spoken in East New Britain, Papua New Guinea; Tagalog (Western Austronesian), spoken in the Philippines; and Inuktitut (Inuit-Yupik-Unangan), spoken in northern Canada. Given the relative uncertainty concerning the universality of typical features of CDL, we examine the corpora for evidence of their presence or absence across our diverse languages. In some places we have drawn upon other data to fill gaps where the sketch corpus could not provide a definitive answer to the presence or absence of a feature, although some gaps still remain and await further study.

4. CDL across five diverse languages

Table 1 outlines the proposed table of contents for an acquisition sketch, zooming in on §5, which deals with CDL. For this section, the manual recommends that sketch writers pay special attention to a number of structural and lexical features that are commonly attested in CDL. We list these features in Table 1, distributed across six subsections and with information on whether a feature was found to be present in our target languages. This recommendation follows from practical considerations: to provide a common starting point and ensure that individual sketches cover similar ground. It is entirely possible that speakers do not modify their language when addressing children, that they only make use of a subset of the features, and/or employ other features. Please see Defina et al. (2023b, pp. 51–63) for further discussions motivating this approach and the choice of features.

Table 1 Investigated features of CDL, their distribution over the (sub-) sections of the acquisition sketch and their presence in the five target languages

Section header & features of CDL	M	*P*	Q	T	I
1. General information					
2. Language typology overview					
3. Data					
4. Learning environment and ethnotheories					
5. Child-directed language					
5.1. Scaffolding techniques					
Many questions, directives, prompts	✓	✓	✓	✓	✓
5.2. Responses to non-target-like utterances	not covered in this contribution				
5.3. Phonology and prosody					
Exaggerated pitch, larger signs	✓	?	✓	✓	?
High F0	✓	?	✓	?	?
Long duration of utterances, long pauses	✓	?	✗	?	?
5.4. Lexicon and semantics					
Reference to the here and now	?	✓	?	✓	✓
Nursery vocabulary	✗	✓	✗	✓	✓
5.5. Morphosyntax					
Short utterances, reduced complexity	?	✓	✓	✓	✓
Few hesitations and errors	✓	✓	✓	✓	✓
Many full and partial repetitions, variation sets	✓	✓	✓	✓	✓
5.6. Gestures	not covered in this contribution				
6. Child language					
7. Community report					

M: Murrinhpatha; *P*: Pitjantjatjara; Q: Qaqet; T: Tagalog; I: Inuktitut
✓: present; ✗: absent; ?: unknown.

Below, we discuss each language of our sample in more detail. In each case, we first introduce the language and dataset, followed by an overview of the features, and end with a summary. As far as possible, the descriptions cover all features and follow roughly the same order. They differ with regard to the level of detail. Each description homes in on one or more focal areas that are especially interesting from the perspective of the particular language and/or which allow us to discuss specific aspects of the sketch approach, while the remaining features are addressed in much less detail. We have deliberately chosen this approach, as it allows us to highlight the richness of the available sketch data.

We note that Barb was a member of the team that collected the Murrinhpatha corpus (see Kelly et al., 2015). We thus begin with CDL in Murrinhpatha.

4.1 Murrinhpatha

4.1.1 Introduction

Murrinhpatha is a non-Pama-Nyungan language of Northern Australia and is part of the Southern Daly language group (Green, 2003). It is spoken by approximately 2,700 people, in the remote Aboriginal community of Wadeye, and the broader Thamarrurr area in the Northern Territory. Murrinhpatha (along with Pitjantjatjara, discussed below) is one of the very few traditional Indigenous languages of Australia that is still acquired by children today and is the language of everyday use in the community (Marmion et al., 2014; National Indigenous Languages Report, 2020).

With the exception of the prosodic analyses, descriptions of Murrinhpatha CDL provided here draw on the Murrinhpatha acquisition sketch corpus. This dataset consists of recordings of two girls, Acacia and Emily, interacting with caregivers and at least one other child, at approximately six-month intervals, from the ages of 2;6/2;7 to 4;3/4;6. These five hours were taken from a longitudinal corpus of 77 h, recorded between 2012 and 2015, following a total of 12 focus children (Davidson, 2018; Forshaw, 2016, 2021). All recordings were made outside at locations of the participants' choosing; for example, a local fishing spot and nearby lookout. For further methodological details of the sketch corpus, see Forshaw et al. (this volume). The data used for the study of prosody in Murrinhpatha CDL do not come from the acquisition sketch. Instead, audio recordings with caregivers as they helped transcribe child–caregiver interactions in the larger corpus are used. Some of the recordings being transcribed make up the Murrinhpatha acquisition sketch corpus.

4.1.2 CDL features

In terms of scaffolding techniques, a common feature of Murrinhpatha CDL is elicitation routines. With children under the age of three years these typically involve the imperatives *thama* 'say/do it', *nana* 'say it to him', or *nange* 'say it to her'. With older children caregivers use these imperative forms as well as display questions (Davidson, 2018). The extent to which caregivers engage children in prompting routines is highly dependent on the individual in question, and the range of caregivers in the sketch corpus reflects this variation. One caregiver was at one end of the spectrum, with roughly half of her utterances to children being prompts. Out of the 108 utterances she directed to Acacia (2;7) in 30 minutes, 55 utterances included one of the three prompting forms listed above, and out of the 146 utterances she directed to Emily (3;2) in 30 minutes, 67 were prompts. Another caregiver in the sketch corpus was at the other extreme. Often quite reticent in general, she rarely engaged her children in elicitation routines. In one 30-minute section of a recording, the two utterances this caregiver directed to her infant (0;11) were prompts (*tata nana* 'Say to him, 'bye!'', 20140416_8.25). In this same recording she directed 17 utterances to her daughter Acacia (3;6) and none of these featured a prompt. Examples (1) and (2) illustrate prompting from the more verbose caregiver:[1]

(1) Caregiver to Acacia (2;7) (20120502_12.46)

kardu	*numi*	*thani-kaykay*	*na-na*
person	other	2SG.BE(4).FUT-call.out	2SG.SAY/DO(34).FUT-3SG.M.IO

'Say to him, 'call out to another person!''

(2) Caregiver to Emily (3;2) (20120502_19.33)

tham-ngarra-werr-warda=thim
2SG.POKE(19).NFUT-1PL.IO-not.give.opportunity-TEMP=2SG.SIT(1).NFUT

thama
2SG.SAY/DO(34).FUT
'Say 'you're not giving us a chance!''

Moving onto phonology and prosody, preliminary work on prosodic features of Murrinhpatha CDL by Jepson and Davidson (2019) described three distinct intonation contours, which correspond to the following functions: "engage", "implore", and "scold". Examples of each melody from at least four different caregivers were analyzed. The "engage" melody is used predominantly – but not exclusively – with children up to the age of around two years, to direct the child's attention to objects and people within their environment. The features of the "engage" contour best align with existing descriptions of CDL prosody in other languages (e.g. Fernald et al., 1989; Kelkar, 1964; Kess & Kess, 1986; Snow, 1972; see Figure 1).

With respect to pitch, both the "engage" and "scold" contours feature a high pitch register and expanded pitch range (~240 Hz and ~425 Hz, respectively). The "implore" contour, in contrast, features a lower pitch register, relatively flat pitch contour, and considerable vowel lengthening. Features of the "engage" contour beyond pitch appear to depend on the individual caregiver, and can include a final rise, constituent-final falsetto, and rounding of lengthened final vowels. In terms of voice quality, palatalization is present in the "engage" melody, but not in the other two. The use of palatalization with infants and its non-use with children over the age of two years is also evident in Murrinhpatha CDL more broadly.

In general, exaggerated pitch appears to be a feature of Murrinhpatha CDL. In the sketch corpus there are numerous examples of higher pitch and extended pitch range used in utterances directed to young children. Also present in the sketch are examples of lengthened vowels in CDL. These are not limited to utterances produced in an "implore" melody, and occur most often immediately prior to the afore-mentioned prompt, *thama* 'say it'.

Looking at lexicon and semantics, caregivers tend to refer to concrete, immediate entities and activities with younger children. For example, looking at one child and her mother, much of the speech directed to Emily at 2;6 relates to the activity that she, her brothers, and her mother are engaged in: constructing houses out of sticks and leaves.

[1]Glossing is in accordance with the Leipzig Glossing Rules with the addition of: ANAPH – anaphora; AV – agent voice; CONJ – conjunction; CONT – continuous aspect; CTG – contingent; DP – discourse particle; ENDR – endearment; IPFV – imperfective; MV – medial verb; NFUT – non-future tense; NPST – non-past tense; PAR – participial; POL – politeness; PV – patient voice; TEMP – temporary.

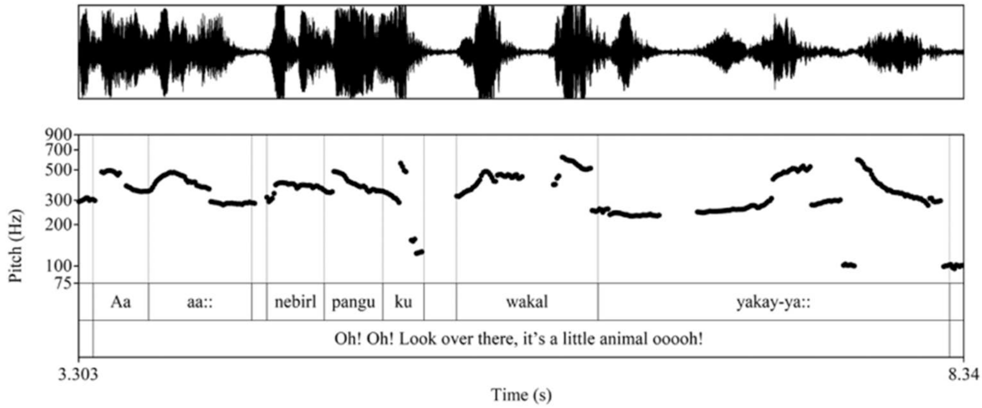

Figure 1 "Engage" contour directed to child (2;0)

Talk to Emily at 3;2 again relates to the "here and now". In contrast, at age 3;8 Emily is scaffolded by her mother to describe a past event involving a wolf (20131105_33.07–35.14); for example, *mampa thangku nekingime thipinhi Mavis nange* 'Tell Mavis what happened to us at night'; *i ku were wulf ngarra mam* 'and what did the wolf do?'.

No examples of nursery vocabulary or modified pronunciations are found in the sketch corpus. This is not surprising as there is no dedicated nursery lexicon in Murrinhpatha CDL, with the exception of the exclamation *anh-anh* 'coochicoo', which is used exclusively with young infants. However, broader observational research in the community suggests some caregivers do adjust their pronunciation of a small number of nouns (but not verbs) when interacting with young children. For example, [ɹ] in *kura* 'water' can be pronounced as [w] (*kuwa*), and the consonant cluster in *biskit* 'biscuit' can be simplified to *bikit*.

With respect to morphosyntax in Murrinhpatha CDL, it is unclear from the sketch data whether caregivers use shorter and/or less complex utterances with young children. This is due to there being insufficient data for comparison. Looking at speech directed to younger children versus older children, it appears that there is little if any difference in the length of utterances and in morphological complexity. Examples (3)–(5) illustrate the speech from one caregiver to her three children of different ages – Emily (2;6) and her two older brothers Bruno (3;8) and Jacob (5;9) – as they engage in the activity of building a little house together:

(3) Mother to Emily (2;6) (20120910_28.49)

mampa	*thurru-tharra*	*ngarra*	*kawu*	*pangu*
alright	2SG.GO(6).FUT-moving	LOC	mother's.mother	DIST

numa-rtarl-nu
2DU.GRAB(9).FUT-tear.off-FUT
'Alright you go over there to your grandma and tear some (paperbark) off.'

(4) Mother to Bruno (3;8) (20120910_27.50)

Bruno	*nhinhika*	*ngarra*	*da*	*rurd*
NAME	2SG-TOP	LOC	PLACE	road

nam-patha=thanam
2SG.HANDS(8).NFUT-make=2SG.BE(4).NFUT
'Bruno where were you making the road?'

(5) Mother to Jacob (5;9) (20120910_29.59)

numi-deyida	*kanhi*	*mampa*	*du-pak*	*kangkarl*
one-same	PROX	alright	2SG.LOWER(17).FUT-put.down	above

'Here's another one, come on, put it on top.'

This observation from the acquisition sketch echoes previous work comparing Murrinhpatha-speaking adults' verb productions in CDL with ADL. It was found that the four caregivers in the study did not modify the length or complexity of the verbs they produced when speaking to young children (Nordlinger et al., 2017).

Minimal hesitations and errors are found in Murrinhpatha CDL. Repetition is commonly used by caregivers, as well as variation sets.

4.1.3 Summary

Prosodic features (e.g. high pitch) and the use of elicitation routines are clear markers of CDL in Murrinhpatha, and it follows that both these features can be found in the sketch. The lack of a nursery lexicon and the minimal, speaker-dependent CDL pronunciations that can be observed in the community are similarly reflected in the sketch corpus. The existence of other potential features of CDL, such as utterance length, however, remain unclear. Two factors contribute to the limited examples of Murrinhpatha CDL in the sketch corpus. The first is that many aspects of CDL in Murrinhpatha are possibly better described as infant direct language (IDL), as they are predominantly used with children under the age of two. Given that the Murrinhpatha sketch begins with children at the ages of 2;6/2;7, this limits the scope to certain features or to the few contexts in which an infant also happens to be present in the recording. The second factor relates to the vast individual variation that is evident with respect to people's use of CDL. The amount of CDL captured in the sketch corpus therefore also depends greatly on the particular caregivers who were present in the recordings.

4.2 Pitjantjatjara

4.2.1 Introduction

Pitjantjatjara is a Pama-Nyungan language spoken by around 3,500 people who refer to themselves as Aṉangu in the land now known as Australia (Australian Bureau of Statistics, 2021). It is the primary language of daily life in the Aṉangu Pitjantjatjara Yankunytjatjara (APY) Lands in remote north-west South Australia and surrounding communities, and children acquire it as their first and often only language until they begin school in English. Child-directed language is a recognized distinct genre in the community and the same term (*iṯi-ku wangka* 'baby talk/talk for baby' or *tjitji-ku wangka* 'child talk/talk for child') is used for both CDL and child language. In conversations with Aṉangu adults about CDL, they have all commented that children up to the age of four or five years do not yet talk like adults and so in order to be better understood, one should mirror the way that they talk. As shown below, particularly in the nursery vocabulary, Pitjantjatjara CDL often appears to be shaped by this intention to mimic child language. The current description is based on the Pitjantjatjara sketch corpus, which consists of five hours of audio and video recordings of five focus children interacting in naturalistic settings with their peers and adults. The corpus, collected between 2016 and 2019, is sampled in a cross-lagged approach with five ages between 2;0 and 4;0 with two children at each age; further details can be found in Wighton et al. (this volume). As with many of the sketch corpora, there was not enough adult-directed language (ADL) to compare against, so comparisons are made between speech directed to the youngest versus the oldest children and any other older children who were present in the recordings.

4.2.2 CDL features

CDL is commonly focused on the here and now with frequent imperatives. Both these patterns are common in Pitjantjatjara CDL as shown in (6) below (see also Wighton, 2021). A particularly high frequency of imperatives is also noted in CDL and child language for the neighbouring related language Warlpiri (Bavin, 1992, 1993).

Pitch is impressionistically higher to children of around two years than it is to children who are four years and older. Pitjantjatjara adults have also commented on and demonstrated exaggerated pitch contours as a stereotypical feature of CDL. Further prosodic investigation of Pitjantjatjara CDL awaits future study.

The use of distinct nursery vocabulary is a particularly distinctive feature of Pitjantjatjara CDL. Nursery forms are based on stereotypical child phonological substitutions such as consonant cluster reduction, initial consonant elision, consonant and vowel harmonization, palatalization of nasals and stops, and gliding of liquids and rhotics. Some examples of these nursery vocabulary adjustments are listed in Table 2. While these substitutions are individually variable in child speech, they are systematized in adult CDL and are listed in the dictionary (Goddard & Defina, 2020). There is individual and contextual variation in whether adults use the nursery vocabulary form, but the form itself is consistent across speakers and contexts. While most nursery forms are clearly derived from the adult form, some do not appear to be, such as *kuku* for *mamu* 'monster' and *ama* for *ipi* 'breastmilk' (though the latter could be related to *mimi* 'breast', changes of /i/ to /a/ are not common outside of harmonization). The nursery vocabulary only replaces a subset of Pitjantjatjara words, particularly those (believed to be) more commonly used with children such as kin terms and things associated with babies (e.g. breastmilk, naps, blankets).

Pitjantjatjara CDL in the sketch corpus was generally error-free with few hesitations. Utterances directed to younger children tended to be shorter than those directed to older children. The average mean length of utterance (MLU) in morphemes of adult utterances directed to two-year-olds was 1.5 compared with 3.3 for utterances directed to four-

Table 2 Pitjantjatjara nursery vocabulary forms

Pitjantjatjara word	English gloss	Adult phonological form	Nursery vocabulary form	Phonological adjustment
iṯi	baby	/ˈiʈi/	/ˈici/	palatalization
kuṯa	older brother	/ˈkuʈa/	/ˈcuca/	palatalization & consonant harmonization
nyara	there	/ˈɲara/	/ˈɲaja/	gliding
pulangkita	blanket	/ˈpulaŋkita/	/ˈpulakita/	cluster reduction
kapi	water	/ˈkapi/	/ˈapa/	initial consonant elision & vowel harmonization
mamu	monster	/ˈmamu/	/ˈkuku/	n/a
ipi	breastmilk	/ˈipi/	/ˈama/	possibly initial consonant elision & i->a from *mimi* 'breast'

year-olds. This reduction in utterance length was often related to omitted arguments and little use of adverbs and adjectives, as can be seen in (6), where a mother talks to her two children Andrew (1;9) and Sander (7;9) (pseudonyms) trying to get them to move away from a branch that she sees as a danger to Andrew. When they do not respond, she scolds them. This example also shows a common simplification noted in Pitjantjatjara CDL where chained imperative verbs are used in place of clause chains with morphologically marked non-finite medial verbs (cf. (6) line iii, which would more commonly in ADL be *a-nkula paka-ṟa* 'move-MV get.up-IMP' or *paka-la ara* 'get.up-MV move-IMP'). This is further discussed in Defina (2020).

(6) Mother to her children Andrew (1;9) and Sander (7;9) (Andrew_20161010_0:14:13)

i. Mother to Sander (7;9)

Sander,	*iṯi*	*pala-ø*	*ngalya-kati*
NAME	baby	that-ACC	this.way-bring.IMP

'Sander, bring that baby (Andrew) this way!'

… [five repeated attempts to get Sander to act]

ii. Mother to Andrew (1;9)

ara
move.IMP
'Move!'

iii. Mother to Andrew (1;9)

ara	*paka-ṟa*
move.IMP	get.up-IMP

'Get up and move!'

iv. Mother to Andrew (1;9)

nyuntu	*pina*	*pati*
2SG.NOM	ear	closed

'Can't you hear?!' (Literally: you have closed ears)

v. Mother to Sander (7;9)

ngana	*Sander*	*pina-ngka*	*kunyu*	*kuli-la*
hey	NAME	ear-LOC	ANAPH	listen-IMP

'Hey Sander, apparently you should listen with your ears!'

Repetition is very common, often in variation sets with small variations between repetitions. This can be seen in (6), lines ii and iii. These variation sets also often feature changes in word order, which is free and highly variable in adult Pitjantjatjara (Wilmoth et al., 2025); for instance, lines i and ii in (7), where the initial statement is in Agent Patient Verb order but is then immediately repeated in Patient Agent Verb order:

(7) Mother to Andrew (2;4) (Andrew_20170506_0:07:56)

i.

wii,	*mami-lu*	*kala*	*rikina-ø*	*palya-ṉi,*	*nya-wa-nta*
hey	mummy-ERG	colour	cool-ACC	make-PRS	look-IMP-DP

'Hey, Mummy is making pretty colours, look.'

ii.

wii,	*kala*	*rikina-ø*	*mami-lu*	*palya-ṉi*
hey	colour	cool-ACC	mummy-ERG	make-PRS

'Hey, Mummy is making pretty colours.'

4.2.3 Summary

In sum, the sketch data suggest that Pitjantjatjara CDL is distinct from ADL in cross-linguistically typical ways, with shorter utterances, some morpho-syntactic simplification, use of variation sets, and nursery vocabulary. The limits of the data mean that any generalizations should be made with caution; however, the patterns observed in the sketch corpus do align with Aṉangu adults' reports that they aim to talk to children as they themselves speak, particularly with the use of higher pitch and nursery vocabulary which features common child phonological substitutions.

4.3 Qaqet

4.3.1 Introduction

Qaqet is a (non-Austronesian) Baining language of Papua New Guinea, spoken by around 15,000 speakers in East New Britain Province (Hellwig, 2019). The language is spoken and acquired in two different settings: as the only or dominant language in the remote mountainous interior, or simultaneously with the dominant language Tok Pisin in the more accessible coastal region. From 2014 to 2018, we compiled longitudinal corpora in both settings, following the children of four families, with a focus on children aged 2;0–4;0 as well as their younger and older siblings (Hellwig et al., 2023b). The children were video-recorded by their parents once a week in natural interaction, resulting in about 400 h of recordings (of which about 100 h are transcribed and annotated). Our sketch corpus of five hours was extracted from the Qaqet-dominant corpus and follows a cross-lagged approach: it comprises data from four children, covering five different age points between them (2;0, 2;6–2;8, 3;1–3;2, 3;7, 4;0). This contribution is based on the sketch data but includes comparisons with analyses from the longitudinal corpus as well as from a controlled study on CDL vs. ADL in narrative contexts (Frye, 2022) in order to explore potentials and limits of the sketch approach.

4.3.2 CDL features

In her controlled study, Frye (2022) investigates 20 retellings of the pear story film (Chafe, 1980), comparing the language of 10 adults in both CDL and ADL contexts. Her study shows that Qaqet CDL systematically differs from ADL in most of the dimensions that are expected from the literature: more attention-organizing utterances such as questions, directives and prompts, higher frequency ranges and higher F0, shorter utterances, and fewer hesitations and errors. Speech rate and pause length did not differ across CDL/ADL, the use of a nursery vocabulary was marginal, and the remaining features (reference to the here and now, repetitions) were not investigated. The sketch data confirm the above observations for conversational contexts and show additionally that repetitions and variation sets are highly frequent in CDL. By way of an example, we highlight the findings on MLU from the sketch corpus.

One characteristic of sketch corpora in general – including also the Qaqet sketch corpus – is that they are unlikely to contain enough ADL to systematically compare CDL and ADL. However, the analysis may reveal differences in the language directed at different age groups, thus allowing us to develop hypotheses on this basis. This is also what we find in the Qaqet sketch data. Table 3 summarizes the MLU (in morphemes) in CDL directed at each of the five target age groups, separated between the CDL of

Table 3 MLU of CDL in the Qaqet sketch corpus (adult and older child speakers)

	Adult speaker		Older child speaker (aged 4;9–8;0)	
Age of target child	No. of CDL intonation units	MLU (morphemes)	No. of CDL intonation units	MLU (morphemes)
2;0	365	3.28	805	2.29
2;6–2;8	76	4.97	282	3.76
3;1–3;2	686	4.69	303	4.17
3;7	19	5.11	332	5.42
4;0	93	4.43	22	4.60

adults and older children. Ethnographic observations suggest that children spend considerable time together with their peers, and this pattern is clearly reflected in the sketch corpus. Given that the corpus features fairly little adult/child interaction, the numbers are aggregated across different contexts, carers, and children.

Table 3 shows an overall increase in MLU with the age of the target child, both for adult speakers and older child speakers. We also observe that the MLU of older child speakers is lower than that of adult speakers when addressing target children in the first three age brackets. A comparison across the age groups is made difficult by the considerable differences in the available data points. However, taking into account only the two age groups with the most data points – target children at age 2,0 and 3;1–3;2 – both interpretations become more robust.

Looking more closely at adult/child interaction, it emerges that large amounts of the data come from a 30-minute stretch where a mother interacts with her two children. That is, the above pattern for adult/child CDL (but not for older child/target child CDL) may reflect individual preferences. On the positive side, the advantage is that it is possible to directly compare the speech of the same adult to children at two different ages: an MLU of 3.61 to the child at age 2;0 vs. an MLU of 4.72 to the child at age 3;2. Furthermore, the contextual factors are identical for both children: the mother is preparing food, and the children are helping clean the vegetables. Examples (8) and (9) illustrate typical utterances and their differing MLUs. In both cases, the mother asks her children to break off usable parts of the vegetables and put them into a dish:

(8) Mother to YDS (2;0) (LongYDS20150506_1 15.588 30.237)

i. *nyi*
'you.'

ii. *nya=tu* *mara*
2SG.F.SBJ=put.CONT here
'you put (it) here.'

iii. *nya=tigis*
2SG.F.SBJ=break.CONT
'you break (it).'

iv. *ma*
'like this.'

v. *ma*
'like this.'

vi. *sa* *ani*
already thither
'here (one more for you).'

(9) Mother to YRA (3;2) (LongYDS20150506_1 145.287 150.763)
i. *nyi=tu-qa=iara* *mer=ama=dis-ki*
2SG.SBJ.NPST=put.CONT-3SG.M=here in=ART=plate-SG.F
'you put it here into the plate.'

ii. *kudi* *nani* *nya=ts-ka*
hopefully can 2SG.SBJ=eat.CONT-3SG.M
'you can eat it later (once the vegetables are cooked).'

iii. *magr=a*
enough=DIST
'enough now.'

iv. *de=nya=tu-nget*
CONJ=2SG.SBJ=put.CONT-3N
'you put them.'

4.3.3 Summary

The pattern in this 30-minute session – as well as the pattern in older child/target child interaction in the whole sketch corpus – shows that at least for some interlocutors and contexts, the MLU increases with age, thus warranting the hypothesis that this pattern may be generalizable beyond the sketch data. Indeed, both the longitudinal corpus and the controlled narrative study confirm this pattern. In particular, Frye (2022, pp. 87–91) shows that CDL is shorter than ADL across individuals (both mothers and fathers) and increases with the age of the child; at age 3;6, CDL/ADL no longer show any difference in MLU.

Given the limits of the sketch data, it is inevitable that not all patterns can be detected in the data (e.g. the observation that adults no longer shorten their utterances when addressing children aged 3;6 and above was not visible in the sketch data), and any generalizability should be approached with caution (e.g. generalizing from the CDL of one mother). That said, the available sketch data allow us to identify and describe patterns that are consistent with the findings from more comprehensive studies. In the absence of additional data sources, the sketch data thus minimally allow us to form hypotheses and identify productive avenues of future research.

4.4 Tagalog

4.4.1 Introduction

Tagalog is a Malayo-Polynesian language from the Philippines with more than 28 million native speakers (Eberhard et al., 2024). We examined Garcia and Kidd's (2022) Tagalog CDL corpus of semi-structured interactions between 20 child–caregiver pairs from Metro Manila, which was collected in 2019 following the acquisition sketch format. The data collection was cross-sectional and the children were between the ages of 2;0 and 4;0. Each pair was recorded at their home for one hour as they used the materials that we provided (e.g. toys, picture books). For the current study, we made use of all the available 20 h of data allowing us to observe certain patterns more clearly (e.g. verb-marking), compared to analyzing only five hours of data.

4.4.2 CDL features

Among utterances with verbs, 29% were imperative or hortative. There were also more than 8,000 occurrences of "?" in the whole CDL corpus, suggesting frequent use of questions. Cursory examination of the transcripts also showed the use of prompts, such as the caregiver asking, 'and then?' to prompt the child to continue telling a story. We also observed productions of incomplete sentences with an intonation that indicated a prompt, such as 'drink a lot of … ?', so the child could complete the sentence.

One of the authors, a native speaker of Tagalog, judged the CDL as having exaggerated pitch contours. Because the recordings had a lot of noise (e.g. electric fan), which also varied per child–caregiver pair, it was difficult to check for F0. Moreover, we did not have ADL from the same adults who provided the CDL, leading to difficulties in comparing the F0 between CDL/ADL. The semi-structured nature of the corpus made it unsuitable for examining the duration of utterances and pauses, as the presence of the toys influenced interaction. Analysis of F0 and pauses thus awaits future research.

In terms of CDL vocabulary, adults in the sketch corpus mostly talked about concepts that were present in the surroundings during the recording. Levshina et al.'s (2023) analysis of 399 CDL utterances with a verb, agent, and patient in Garcia and Kidd's (2022) corpus showed that only 3% of agents and only 5% of patients were not present in the environment. Levshina et al. collected Tagalog ADL corpus data, and found that in 416 ADL utterances with a verb, agent, and patient, 27% of agents and 81% of patients were not present or were non-referential (e.g. "one" in 'One has to follow the rules'). However, the ADL data were from audio recordings while the CDL data were from videos, so the latter was easier to judge than the former in determining if a referent was present or not. CDL also showed the use of nursery vocabulary such as *mamam* for 'water', *teddy* for 'bear', and *tweet-tweet* for 'bird'. Another feature of Tagalog CDL seems to be the dropping of noun markers (2% for the first noun phrase, 7% for the second noun phrase).

We found that CDL was characterized by short utterances, with an MLU in words of 3.09 (range: 2.40–3.92), as calculated using CLAN (MacWhinney, 2000). This average length is shorter than the mean we found in Levshina et al.'s (2023) Tagalog ADL corpus, which was 5.28 (range: 4.26–6.52). Only 0.1% of all the words in the CDL corpus were marked as having an error, which typically included an incorrect verb inflection or an error in noun marking. There were also many instances of full and partial repetitions and variation sets, especially because verbs are not only inflected for aspect and mode in the language, but also for voice, which we turn to next.

Tagalog has a unique voice-marking system where the inflection on the verb dictates the thematic role of the noun phrase that is marked by *ang*, which is considered as the subject (Himmelmann, 2005). For instance, as illustrated in (10), the agent voice (AV) infix *-um-* assigns the *ang*-phrase the agent role. In contrast, as shown in (11), the patient voice (PV) infix *-in-* denotes that the *ang*-phrase is the patient. Hence, the roles of agent and patient are reversed with only a change in the voice-marking on the verb. The thematic role assignments are not affected by the order of the arguments such that there is no change in meaning between agent voice agent-initial and patient-initial (10) sentences, and the same is true for the patient voice.

(10)	S<um>i~sipa	ng	baboy	ang	baka
	<AV>IPFV~kick		pig		cow
	'The cow is kicking a pig.'				

(11)	S<in>i~sipa	ng	baboy	ang	baka
	<PV>IPFV~kick		pig		cow
	'The/A pig is kicking the cow.'				

Out of the 9,401 CDL utterances with verbs, 45% of verbs were marked for the patient voice, 21% for the agent voice, and 34% of verbs were not inflected for voice. This lack of marking seems to be a characteristic of CDL, although the frequency of dropping the markers did not depend on the child's language skills (as measured by MLU in morphemes), but rather mostly on the individual caregivers (Garcia & Kidd, 2022). Of the 2,504 utterances with uninflected verbs but with at least one marked noun, 66% should have been in the patient voice based on the noun-marking. This finding means that CDL predominantly exhibits the patient voice mapping, such that the subject is commonly not an agent. In terms of word order, out of the 7,960 utterances with at least one noun, 73% were agent-initial or agent-only. Splitting by voice, 86% of utterances with agent voice verbs, 63% of utterances with patient voice verbs, and 82% of non-voice-inflected verbs were agent-initial or agent-only. Overall, regardless of voice, the predominant word order in CDL is agent first. Thus, despite the potential complexity associated with voice marking and flexible word order, Tagalog CDL is somewhat simplified, which may function to help children bootstrap into this aspect of the language.

4.4.3 Summary

Overall, similar to patterns found in other languages, Tagalog CDL differs from ADL as the former is typically shorter, with exaggerated pitch contours, more references to the here and now, and is characterized by the use of nursery vocabulary, and frequent use of questions, imperatives, repetitions, and variation sets. We also found that CDL was modulated by individual variation. These varying characteristics are better captured given more available data.

4.5 Inuktitut

4.5.1 Introduction

Inuktitut is a polysynthetic and agglutinative language of the Inuit-Yupik-Unangan family, spoken by around 40,000 speakers in the eastern arctic region of Canada (Statistics Canada, 2023). The data discussed here come from the Nunavik region (northern Quebec), where about 98% of Inuit children learn Inuktitut as their first language and Inuktitut is the predominant language spoken in the community (Statistics Canada, 2023). The data are taken from two larger longitudinal studies conducted in two communities of 300–400 inhabitants – one on communicative competence (four children aged 1;0–2;9, 64 h recorded in 1986–1987, 32 h transcribed and annotated; Crago, 1988) and the other on the acquisition of morphosyntax (four children aged 2;0–3;6, 144 h recorded in 1989, 72 h transcribed and annotated; Allen, 1996). The children were all videotaped by researchers in naturalistic interactions with peers and caregivers. The sketch corpus was extracted from these two larger corpora and comprises 30 minutes of data from each of two children at ages 1;4, 1;10, 2;4, 2;10, and 3;4 (Lee & Allen, 2023). Only mothers' utterances are considered in the analyses of CDL; unfortunately, no comparison data from ADL were available. Findings from the sketch corpus, as published in Lee and Allen (2023), are compared with findings from (subsets of) the larger corpus when available.

4.5.2 CDL features

Scaffolding techniques typical of CDL were very evident in Inuktitut. Interactive routines for modelling speech were common. Several interactions in the sketch corpus featured greeting routines and modelling correct speech using *la-* 'say … ' or *imaak* 'like this', both patterns found consistently in the larger corpus (Crago, 1988) and illustrated in (12):

(12) Mother to Lizzie (2;6) (Lee & Allen, 2023, p. 153)
Saali-ai *la-git.*
Charlie-hi say-IMP.2SG
'Say "hi Charlie".'

Speech acts typical of CDL were also common. Imperatives were very frequent in the sketch data, comprising between 15% and 33% of utterances across the ages with no clear developmental change. They were typically used to instruct behaviour and speech, and to offer help to the child. Questions were also commonly used, predominantly to ask for information and to check comprehension. Mothers did not typically correct children's non-target-like utterances in the sketch data. When they did so, they typically reframed the utterance or pointed out the non-target-like features to another interlocutor.

In terms of the phonetics and phonology of utterances in CDL, we have less information. None of the typical features mentioned in the CDL literature – slower speech rate, exaggeration of pitch contours, F0, longer speech duration, or more pauses – are mentioned in cultural descriptions of CDL in Inuktitut or in previous research (e.g. Crago, 1988; Ekho & Ottokie, 2000). Since we have no sample of ADL, we were not able to compare CDL vs. ADL for any of these measures. We did measure speech rate in the sketch corpus; mothers used around six syllables per second on average but showed no noticeable pattern of change by age.

For vocabulary, we found clearer patterns. Inuktitut has a well-established nursery vocabulary that is phonologically simpler and shorter than the corresponding adult words, often including reduplicated syllables or onomatopoeia; for example, *apaapa* vs. *niqi* 'food', *maa-* vs. *kuni-* 'kiss', *uppa-* vs. *atsuruq-* 'lift/pick up'. These baby words comprised between 7% and 20% of noun and verb roots at 1;4 and 1;10 in the sketch corpus but did not appear after that, although data from a larger corpus show that they are used by and to children up to age 3;0 (Crago, 1988; Crago & Allen, 1997). Mothers also used a wider range of vocabulary in the sketch data as the children aged, with later-used semantic categories including furniture, vehicles, and places to go. This increase in lexical diversity of CDL with child age is also consistent with patterns found in a larger corpus (Johnson & Allen, 2025).

In the area of morphosyntax, we found that utterances in Inuktitut CDL tend to be relatively short and complete with relatively few hesitations and errors, although this has not

Table 4 Average MLU in morphemes in child-directed language according to age of child (note that Mother A and Mother B represent the two mothers at a given age, and are not necessarily the same person across ages)

	1;4	1;10	2;4	2;10	3;4
Mother A	2.35	2.65	3.50	4.62	4.52
Mother B	2.50	2.86	3.22	3.62	4.28
Average	2.43	2.76	3.36	4.12	4.40

been directly compared with ADL. The mean length of utterance in morphemes increased between speech directed to younger vs. older children in the sketch corpus, as shown in Table 4 (Lee & Allen, 2023, p. 163).

Interestingly, the typical length of words (in morphemes) also increased with the child's age, as one might expect in an agglutinative and polysynthetic language. At 1;4, for example, 42% of words had one morpheme, 3% had four morphemes, and 0% had seven morphemes. At 3;4, in contrast, 30% of words had one morpheme, 16% had four morphemes, and 2% had seven morphemes. This is consistent with patterns based on a larger corpus (Johnson & Allen, 2022), and suggests that the mothers are gradually introducing the children to more complex word structure as they get older and as their linguistic ability increases.

Other aspects of morphosyntax also showed changes with child age in the sketch corpus. Mothers used more verbal and nominal inflections as the children aged, and also a greater variety of each. The earliest and most common verbal inflections used were imperatives, while inflections in the conditional and dubitative moods only appeared at the older ages; this is consistent with the pattern found in a larger corpus (Lee et al., 2023). The earliest and most common nominal inflections were absolutive and modalis singular forms including possession, while vialis inflections appeared much later and no equalis inflections were used. Mothers also occasionally used variation sets to highlight the inflections, as exemplified in (13) for verbal inflections (*-li* vs. *-git* on stem *amaaliq-* 'carry piggyback') and (14) for nominal inflections (*-nga* vs. no inflection on the root *angajuapik* 'elder sibling'):

(13) Verbal inflection: Mother to Sarah (1;4) (Lee & Allen, 2023, p. 167)

i. *ma-ani* *amaa-lir-li.*
here-LOC carry.piggyback-POL-IMP.3SG
'Let it be carried on your back here.'

ii. *atii* *amaa-li-rit.*
initiate carry.piggyback-POL-IMP.2SG
'Come on, carry it on your back.'

(14) Nominal inflection: Mother to Jini (1;4) (Lee & Allen, 2023, p. 171)

i. *angaju-api-nga=li?*
elder.same.sex.sibling-ENDR-POSS.3SG>3SG.ABS=where
'Where's his/her older sister?'

ii. *angajuk-apik* *imaak.*
elder.same.sex.sibling-ENDR like.this
'(Say) older sister like that.'

Two other forms – demonstratives and verbalizers – showed more and more varied use in CDL as children aged, consistent with patterns in a larger corpus (Johnson & Allen, 2022; Lee & Allen, in press). For demonstratives, the earliest and most frequent forms were the proximal static *una* 'this one' and *ukua* 'these ones'. Later forms included the distal static (*inna* 'that one', *ikkua* 'those ones') and dynamic forms (*taanna* 'that one', *taakkua* 'those ones'). For verbalizers, the earliest and most frequent forms were *-it-* 'be' and *-tuq-* 'consume' as in (15a), while later forms included *-guq-* 'become' and *-qaq-* 'have' as in (15b):

(15) a. Mother to Jini (1;4) (Lee & Allen, 2023, p. 172)
tii-tu-langa-junga.
tea-consume-FUT-PAR.1SG
'I'll have some tea.'

b. Mother to Lizzie (2;10) (Lee & Allen, 2023, p. 153)
nani *qimmi-qa-nngi-mmat?*
where.at dog-have-NEG-CTG.3SG
'Where are there no dogs?'

In contrast, the percentage of null subjects (89–98%) and null objects (69–86%) showed no clear variation in CDL with child age. Valency-alternating structures such as passive and causative were used in CDL across all ages, but not frequently enough to see any change with age.

4.5.3 Summary

In sum, child-directed language in Inuktitut shows many similarities with patterns found in other languages – use of socialization routines, frequent imperatives and questions, nursery vocabulary, short and complete utterances, and use of variation sets. Our research also showed that mothers are sensitive to their children's linguistic level in that they increase the variety of vocabulary as well as the complexity of morpheme and word structure as the children's linguistic ability increases – a pattern also seen for other structures in a larger corpus (Johnson & Allen, 2022; Lee et al., 2023).

5. Discussion

Our analysis of the distribution of CDL features across five diverse cultures suggests the presence of recognizable but variable registers in each language, as one might expect given the variable nature of CDL cross-linguistically. Some features were observable in all five languages, whereas other features were either absent or infrequent enough in some languages to not be observed in the sketch corpus. Overall, while there was variation, Table 1 is notable for the fact that most cells contain a positive observation of the feature (i.e. more cells are marked with a check than an "x"). This suggests two immediate conclusions. Firstly, the features are good candidates as *likely* features of CDL for any one language. Secondly, the sketch approach is well suited to the study of CDL.

All corpora revealed evidence that CDL had few hesitations and errors, had a large proportion of questions, imperatives, and prompts, and contained many repetitions and variation sets. These features reveal a tendency to provide clear and engaging language to the child across cultures. Caregivers are regularly ensuring that their speech to young children is clear and in most cases error-free, and via the use of repetition and variation sets, they are implicitly providing important distributional information to children about the use of morphemes and words in the language. Whether these modifications serve to help children learn the language is a question for further study, and we note that the evidence is equivocal on this matter (see Kemp et al., 2024). However, it is not out of the question to suggest that these common features may have some pathbreaking role. Children are expert distributional learners (Braine, 1987; Saffran et al., 1996), and thus carefully articulated and repetitious but immediately variable input may ease children's path into the complexities of the language. The use of questions, imperatives, and prompts varies across individuals and with age, and it is notable that in our five languages imperatives are commonly used with young children. This is a common observation outside of European cultures (e.g. Abels et al., 2021). All three speech acts serve to regulate the child's behaviour within interaction, whether it is for behavioural control, as in imperatives, or to draw them into the conversation, as in prompts and questions. The structured and child-centred nature of these behaviours is consistent with socio-cultural approaches to child development (e.g. Bruner, 1983; Nelson, 1996; Tomasello, 1999; Vygotsky, 1978), whereby the development of cognitive processes like language is argued to be scaffolded by the behaviours of competent others.

While there are commonalities across the five languages, there is still clear variation both across languages and, importantly, amongst speakers within languages. For instance, there is substantial use of a nursery vocabulary in Pitjantjatjara and Inuktitut, and little evidence of it in Murrinhpatha and Qaqet. In the two polysynthetic languages in our sample, Murrinhpatha and Inuktitut, there was simplification of utterance length according to the age of the child in Inuktitut only, with Murrinhpatha caregivers notably not simplifying the functionally important verb complex. This lack of simplification is again contrasted with another morphologically rich language, Tagalog, where caregivers often produced verbs unmarked for voice. This is ungrammatical in the adult language, but in CDL may serve to alert children to how verbs are marked for voice via infixes while at the same time maintaining the more frequent patient voice noun marking pattern.

One common observation from all languages was that CDL varied across individual caregivers. Given the small samples that a typical sketch corpus deals with, this is not overly surprising, and it is difficult to know whether the differences were stable properties of individuals, context-specific reactions to specific recordings, or a combination of both. However, individual variability was also evident in voice marking in Tagalog, which had a larger sample of caregivers ($N = 20$); the best predictor of whether a verb would be marked for voice was the individual caregiver (see Garcia & Kidd, 2022). This raises a larger point about CDL: although it has historically most often been referred to as a register, it is really a dynamic sociolinguistic phenomenon that caregivers flexibly use in a context-specific manner (for discussion see Weinstein & Baldwin, 2024). Under this categorization, we should expect variation in its use within and across individuals in a given linguistic community.

The discussion of variability brings us back to the second point: the suitability of the sketch approach to the study of CDL. While variability in the sketch data cannot be straightforwardly attributed to specific factors, it is interesting to observe that variability is detectable in small amounts of data, and also that it does not get in the way of identifying and describing patterns. More generally, it is clear that the sketch data only capture a snapshot of the patterns and possibilities available in a language, and thus any generalizations need to be calibrated accordingly. However, in quite a few cases, we were able to compare this snapshot with studies based on larger datasets and/or more controlled studies. In all cases, the findings were consistent. For low frequency phenomena, it is possible that they do not show up in the sketch data. For example, the Murrinhpatha and Qaqet sketch corpora did not include any nursery vocabulary items or modified pronunciations. Other sources reveal that a few such items and pronunciations do in fact exist in both languages, but given their low frequency, the sketch data did not capture them. Highly frequent phenomena, by contrast, feature prominently in the sketch data. Thus, (cautious) generalizations are indeed possible in many cases. A particular strength of the sketch approach is its focus on the family/household setting and the concomitant shift away from dyadic interaction. As a result, the corpora feature an abundance of parallel examples where the same adult is talking to children of different ages, thus offering the opportunity for contrastive analyses of language directed at children of different ages (cf. especially the discussion of (3)–(5) and (8)–(9)). Furthermore, the coverage across five age points (2;0–4;0) mitigates a potential drawback of the sketch approach: the limited amount of ADL in the sketch data. Comparing the language addressed to younger vs. older children makes it possible to identify differences and similarities, and thus to formulate plausible hypotheses about the presence or absence of a specific feature of CDL, which can then form the basis for future in-depth comparisons with ADL.

6. Conclusion

In this paper we have presented an analysis of child-directed language in five diverse languages using the *Acquisition Sketch* approach. The overall impression is that, despite the diversity in cultures, there were clear commonalities in how children are spoken to. At the same time, no one language was observed to use an identical set of features in our set, reinforcing our point that CDL as a behaviour is best conceived in terms of family resemblance. Throughout this contribution, our focus was on delineating the distribution of a selected set of features across five languages, while at the same time testifying to the wealth and beauty of the sketch data. Obviously, these feature inventories can only be a starting point. As they stand, they provide plenty of food for thought, as they point out some unexpected patterns that warrant future in-depth studies (e.g. the frequency of imperatives in Inuktitut does not seem to change with the age of the child; we do not observe reduced morphosyntactic complexity in Murrinhpatha). In particular, they also offer the opportunity to probe deeper into some behaviours and systematically compare them across corpora (of related languages, neighbouring languages, typologically similar languages or diverse languages), e.g. comparing the functions of repetitions and variation sets across languages, etc. There is much more cross-cultural work to do on this topic, and the sketch approach is designed to enable these kinds of comparisons without the need for prohibitively large amounts of data collection, caveats on interpretation notwithstanding. As more sketches are added, we will move closer to a more representative understanding of the conditions under which children learn language, which, amongst many giants of the field, will be part of Barb's legacy.

Acknowledgements

We are deeply indebted to Barb, our dear friend and colleague. Barb worked with us developing the Sketch Acquisition Manual (which she kept calling "SAM" until it stuck), and words cannot describe the sadness we felt in her not seeing the final version of it. In another reality, she would be an author on this paper. We are also indebted to the children, caregivers, and communities who contributed their time to our respective projects. Many thanks to two anonymous reviewers for their thoughtful engagement with an earlier version of this paper. Order of authorship is reverse alphabetical.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The Acquisition Sketch Project was supported by a research travel grant from the DAAD/Universities Australia funding scheme (Project-ID 57446167). Research on Murrinhpatha was supported by the ARC (DP110100961). Research on Pitjantjatjara was supported by the ARC (CE140100041). Research on Qaqet was supported by the ARC (FT0991412), ELDP (SG0110), and the Volkswagen Foundation (Az87100). Research on Tagalog was supported by the Max Planck Institute for Psycholinguistics. Research on Inuktitut was supported by Kativik Ilisarniliriniq - the School Board of Nunavik, and by the Social Sciences and Humanities Research Council of Canada.

Data availability statement

Please contact the corresponding author, EK, for details on how to access the sketch corpora.

Notes on contributors

Evan Kidd is currently a Professor of Linguistics at The Australian National University. His research focuses on psycholinguistics, predominantly on child language acquisition and adult language processing, and has an explicit cross-linguistic focus. His current research includes a focus on how typological variables influence online language comprehension and sentence production in both children and adults, cross-linguistic generalizability in psycholinguistic theory, and the neurocognitive learning mechanisms underlying language.

Birgit Hellwig is a Professor of Linguistics at the University of Cologne, where she combines language documentation with language acquisition and socialization research. She is currently working with the Qaqet in Papua New Guinea, and she continues to be interested in the documentation and description of the adult language, researching Goemai (a Chadic language of Nigeria), Katla (a Niger-Congo language of Sudan), and Tabaq (a Nilo-Saharan language of Sudan).

Rowena Garcia is a postdoctoral researcher in the Language Development and Multilingualism area of the Leibniz-Centre General Linguistics (ZAS) in Berlin, Germany, and a visiting professor at the Speech Pathology Department of the University of the Philippines Manila. Her main interests include investigating children's language acquisition and processing, as well as developing language assessment tools, with a particular focus on understudied languages such as Tagalog. She uses corpus data analyses and (field) psycholinguistic methods like eye-tracking.

Rebecca Defina is a linguist at the University of Melbourne. Her primary interests include child language acquisition, Australian Indigenous languages, and event cognition.

Lucinda Davidson is a postdoctoral researcher at the Research Unit of Indigenous Language, at the University of Melbourne. Lucy's research explores multiple aspects of children's language development, with a current focus on multimodal interaction amongst Murrinhpatha-speaking families.

Shanley Allen is Professor of Psycholinguistics and Language Development at the University of Kaiserslautern-Landau. Her research focuses on the development of morphosyntax, with a particular interest in how cross-linguistic differences influence language development. She has published extensively on the development of argument structure and other morphosyntactic phenomena in Inuktitut, on the role of discourse-pragmatics in children's referential choice cross-linguistically, and on language processing in L1 and bilingual children and adults. She has also published on the relation between gesture and speech in describing motion events, the role of child-directed speech in language development, and the structure of code mixing.

ORCID

Evan Kidd http://orcid.org/0000-0003-4301-2290
Birgit Hellwig http://orcid.org/0000-0003-4696-0827
Rowena Garcia http://orcid.org/0000-0003-1363-542X
Rebecca Defina http://orcid.org/0000-0002-3637-1271
Lucinda Davidson http://orcid.org/0000-0001-5175-7394
Shanley Allen http://orcid.org/0000-0002-5421-6750

References

Abels, M., Kilale, A., & Vogt, P. (2021). Speech acts addressed to Hazda infants in Tanzania: Cross-cultural comparison, speaker age, and camp livelihood. *First Language*, *41*(3), 294–313. https://doi.org/10.1177/0142723720972000

Allen, S. E. M. (1996). *Aspects of argument structure acquisition in Inuktitut*. John Benjamins.

Australian Bureau of Statistics. (2021). *Census data summary: Language spoken at home by sex (LGA)*. Retrieved August 7, 2024, from https://stat.data.abs.gov.au/Index.aspx?DataSetCode = ABS_C16_T09_LGA.

Bavin, E. L. (1992). The acquisition of Warlpiri. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition* (pp. 309–372). Erlbaum.

Bavin, E. L. (1993). Language and culture: Socialisation in a Warlpiri community. In M. Walsh, & C. Yallop (Eds.), *Language and culture in aboriginal Australia* (pp. 85–96). Aboriginal Studies Press.

Bernstein Ratner, N., & Pye, C. (1984). Higher pitch in BT is not universal: Acoustic evidence from Quiché Mayan. *Journal of Child Language*, *11*(3), 515–522. https://doi.org/10.1017/S0305000900005924

Braine, M. D. S. (1987). What is learned in acquiring word classes: A step toward an acquisition theory. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 65–87). Erlbaum.

Bromham, L., Dinnage, R., Skirgård, H., Ritchie, A., Cardillo, M., Meakins, F., Greenhill, S., & Hua, X. (2022). Global predictors of language endangerment and the future of linguistic diversity. *Nature Ecology & Evolution*, *6*(2), 163–173. https://doi.org/10.1038/s41559-021-01604-y

Bruner, J. S. (1983). *Child's talk: Learning to use language*. Oxford University Press.

Casillas, M., Brown, P., & Levinson, S. C. (2020). Early language experience in a Tseltal Mayan village. *Child Development*, *91*(5), 1819–1835. https://doi.org/10.1111/cdev.13349

Chafe, W. L. (1980). *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Ablex.

Crago, M. B. (1988). *Cultural context in communicative interaction of Inuit children* [Unpublished doctoral dissertation]. McGill University, Montreal, Canada.

Crago, M. B., & Allen, S. E. M. (1997). Linguistic and cultural aspects of simplicity and complexity in Inuktitut (Eskimo) child-directed speech. In E. Hughes, M. Hughes, & A. Greenhill (Eds.), *Proceedings of the 21st annual Boston university conference on language development* (pp. 91–102). Cascadilla Press.

Davidson, L. (2018). *Allies and Adversaries: Categories in Murrinhpatha speaking children's talk* [Unpublished doctoral dissertation]. University of Melbourne, Australia.

Davidson, L., & Kelly, B. F. (2021). The pragmatics of managing children's distress in Murrinhpatha, a traditional Australian language. *Journal of Pragmatics*, *184*, 167–184. https://doi.org/10.1016/j.pragma.2021.08.008

Defina, R. (2020). Acquisition of Pitjantjatjara clause chains. *Frontiers in Psychology*, *11*, 541. https://doi.org/10.3389/fpsyg.2020.00541

Defina, R., Allen, S. E. M., Davidson, L., Hellwig, B., Kelly, B. F., & Kidd, E. (2023a). Sketch acquisition manual (SAM), part I: The sketch corpus. *Language Documentation and Conservation Special Publication*, *28*, 5–38. https://hdl.handle.net/10125/74719.

Defina, R., Allen, S. E. M., Davidson, L., Hellwig, B., Kelly, B. F., & Kidd, E. (2023b). Sketch acquisition manual (SAM), part II: The acquisition sketch. *Language Documentation and Conservation Special Publication*, *28*, 39–86. https://hdl.handle.net/10125/74720.

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (2024). *Ethnologue: Languages of the world* (27th ed.). SIL International. http://www.ethnologue.com.

Ekho, N., & Ottokie, U. (2000). *Childrearing practices (Vol. 3 of Interviewing Inuit Elders)*. Nunavut Arctic College.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance to cognitive science. *Behavioral and Brain Sciences*, *32*(5), 429–492. https://doi.org/10.1017/S0140525X0999094X

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, *16*(3), 477–501. https://doi.org/10.1017/S0305000900010679

Forshaw, W. (2016). *Little kids, big verbs: The acquisition of Murrinhpatha bipartite verb stems* [Unpublished doctoral dissertation]. University of Melbourne, Australia.

Forshaw, W. (2021). *The acquisition of complex morphology: Insights from Murrinhpatha*. John Benjamins.

Frank, M. C., Alcock, K. J., Arias-Trejo, N., Aschersleben, G., Baldwin, D., Barbu, S., Bergelson, E., Bergmann, C., Black, A. K., Blything, R., Böhland, M. P., Bolitho, P., Borovsky, A., Brady, S. M., Braun, B., Brown, A., Byers-Heinlein, K., Campbell, L. E., Cashon, C., … Höhle, B. (2020). Quantifying sources of variability in infancy research using the infant-directed speech preference. *Advances in Methods and Practices in Psychological Science*, *3*(1), 24–52. https://doi.org/10.1177/2515245919900809

Frye, H. (2022). *Child-directed speech in Qaqet: A language of East New Britain*. ANU Press.

Gallaway, C., & Richards, B. J. (1994). *Input and interaction in language acquisition*. Cambridge University Press.

Garcia, R., & Kidd, E. (2022). Acquiring verb-argument structure in Tagalog: A multivariate corpus analysis of caregiver and child speech. *Linguistics*, *60*(6), 1855–1906. https://doi.org/10.1515/ling-2021-0107

Goddard, C., & Defina, R. (2020). *Pitjantjatjara/Yankunytjatjara to English dictionary* (2nd updated and rev. ed.). IAD Press.

Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby)talk to me: The social context of infant-directed speech and its effects on early language acquisition. *Current Directions in Psychological Science*, *24*(5), 339–334. https://doi.org/10.1177/0963721415595345

Green, I. (2003). The genetic status of Murrinh-patha. In N. Evans (Ed.), *The Non-Pama-Nyungan languages of Northern Australia: Comparative studies of the continent's most linguistically complex region* (pp. 159–184). Pacific Linguistics.

Hellwig, B. (2019). *A grammar of Qaqet*. De Gruyter Mouton.

Hellwig, B., Allen, S. E. M., Davidson, L., Defina, R., Kelly, B. F., & Kidd, E. (2023a). *The acquisition sketch project*. (Language Documentation & Conservation, Special Publication 28.) Honolulu: University of Hawai'i Press. https://nflrc.hawaii.edu/ldc/sp28/.

Hellwig, B., Dawuda, C., Frye, H., & Reetz, S. (2023b). *Qaqet child language corpus*. Data Center for the Humanities.

Himmelmann, N. P. (2005). Tagalog. In A. Adelaar, & N. P. Himmelmann (Eds.), *The Austronesian languages of Asia and Madagascar* (pp. 350–376). Routledge.

Ingram, D. (1995). The cultural basis of prosodic modifications to infants and children: A response to Fernald's universalist theory. *Journal of Child Language*, *22*(1), 223–233. https://doi.org/10.1017/S0305000900009715

Jepson, K., & Davidson, L. (2019). *The highs and lows of childcare: Mapping caregiver melodies in Murrinhpatha* [Paper presentation]. Australian Languages Workshop, Marysville, VIC.

Johnson, O. A., & Allen, S. E. M. (2022). The use of complex structures with a word class change in Inuktitut child-directed speech. *Frontiers in Psychology*, *13*, 971395. https://doi.org/10.3389/fpsyg.2022.971395

Johnson, O. A., & Allen, S. E. M. (2025). Measuring lexical diversity in Inuktitut. Manuscript submitted for publication.

Kelkar, A. R. (1964). Marathi baby talk. *Word*, *20*(1), 40–54. https://doi.org/10.1080/00437956.1964.11659812

Kelly, B. F. (2006). The development of construction through early gesture use. In E. V. Clark, & B. F. Kelly (Eds.), *Constructions in acquisition* (pp. 15–30). CSLI Publications.

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. https://doi.org/10.1177/0142723715602350

Kelly, B. F., & Nordlinger, R. (2014). Fieldwork and first language acquisition. In L. Gawne, & J. Vaughn (Eds.), *Selected papers from the 44th conference of the Australian linguistics society* (pp. 178–192). Australian Linguistics Society. http://bit.ly/ALS2013Proceedings.

Kemp, V., Ota, M., & Schaeffler, S. (2024). Does child-directed speech facilitate language development in all domains? A study space analysis of the existing evidence. *Developmental Review*, *72*, 101121. https://doi.org/10.1016/j.dr.2024.101121

Kess, J. F., & Kess, A. C. (1986). On Nootka baby talk. *International Journal of American Linguistics*, *52*(3), 201–211. https://doi.org/10.1086/466018. http://www.jstor.org/stable/1265174.

Kidd, E., & Garcia, R. (2022). How diverse is child language acquisition research? *First Language*, *42*(6), 703–735. https://doi.org/10.1177/01427237211066405

Küntay, A., & Slobin, D. I. (1996). Listening to a Turkish mother: Some puzzles for acquisition. In D. I. Slobin, J. Gerhardt, A. Kyratzis, & J. Guo (Eds.), *Social interaction, social context, and language: Essays in honor of Susan Ervin-Tripp* (pp. 265–286). Erlbaum.

Lee, H., & Allen, S. E. M. (2023). An acquisition sketch of Inuktitut. *Language Documentation and Conservation Special Publication*, *28*, 135–213. https://hdl.handle.net/10125/74723.

Lee, H., & Allen, S. E. M. (in press). The acquisition of demonstratives and locatives in Inuktitut. *Journal of Child Language*.

Lee, H., Johnson, O. A., & Allen, S. E. M. (2023). The use of verbal inflections in Inuktitut child and child-directed speech. *Journal of Monolingual and Bilingual Speech*, *5*(1), 29–58. https://doi.org/10.1558/jmbs.23491

Levshina, N., Garcia, R., Skyra, Y., & Petukhova, A. (2023). *Given before new? New insights from conversational corpora* [Paper presentation]. 56th Annual Meeting of the Societas Linguistica Europaea (SLE), Athens, Greece.

Lieven, E. V. M. (1994). Crosslinguistic and crosscultural aspects of speech addressed to children. In C. Gallaway, & B. J. Richards (Eds.), *Input and interaction in child language acquisition* (pp. 56–73). Cambridge University Press.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Erlbaum.

Marmion, D., Obata, K., & Troy, J. (2014). *Community, identity, wellbeing: The report of the second national indigenous languages survey*. Australian Institute of Aboriginal and Torres Strait Islander Studies.

National Indigenous Languages Report. (2020). Department of Infrastructure, Transport, Regional Development, Communication and the Arts, Government of Australia. https://www.arts.gov.au/publications/national-indigenous-languages-report-document.

Nelson, K. (1996). *Language in cognitive development: The emergence of the mediated mind*. Cambridge University Press.

Nordlinger, R., Forshaw, W., & Kelly, B. F. (2017). *Learning complex predicates: The role of input* [Paper presentation]. Congress of the International Association for the Study of Child Language, Lyon, France.

Ochs, E. (1985). Variation and error: A sociolinguistic approach to language acquisition in Samoa. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition* (pp. 783–838). Erlbaum.

Ochs, E., & Schieffelin, B. (1984). Language acquisition and socialization: Three developmental stories. In R. Shweder, & R. LeVine (Eds.), *Culture theory: Mind, self, and emotion* (pp. 227–320). Cambridge University Press.

Passmore, S., Hellwig, B., Garcia, R., & Kidd, E. (2025). The scientific and cultural cost of convenience sampling in the face of rising language endangerment: Highlighting the role of language acquisition. *Open Mind: Discoveries in Cognitive Science*, (9), 501–514. https://doi.org/10.1162/opmi_a_00199

Pye, C. (1986). An ethnography of Mayan speech to children. *Working Papers in Child Language*, *1*, 30–58.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Schieffelin, B. (1985). The acquisition of Kaluli. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition* (pp. 525–593). Erlbaum.

Snow, C. E. (1972). Mothers' speech to children learning language. *Child Development*, *43*(2), 549–565. https://doi.org/10.2307/1127555

Snow, C. E., & Ferguson, C. A. (1977). *Talking to children: Language input and acquisition*. Cambridge University Press.

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, *27*(4), 501–532. https://doi.org/10.1016/j.dr.2007.06.002

Statistics Canada. (2023). *Indigenous languages across Canada*. Retrieved September 15, 2024, from https://www12.statcan.gc.ca/census-recensement/2021/as-sa/98-200-X/2021012/98-200-X2021012-eng.cfm.

Tomasello, M. (1999). *The cultural origins of human cognition*. Harvard University Press.

Vygotsky, L. S. (1978). *Mind in society: The development of higher mental process*. Harvard University Press.

Weinstein, N., & Baldwin, D. (2024). Reification of infant-directed speech? Exploring assumptions shaping infant-directed speech research. *Culture and Psychology*, *30*(1), 216–242. https://doi.org/10.1177/1354067X221147683

Wighton, W. (2021). *Children's verbal inflection development in Pitjantjatjara: An acquisition sketch* [Graduate Diploma in Arts (Advanced) dissertation]. University of Melbourne, Australia. http://hdl.handle.net/11343/290224.

Wilmoth, S., Nordlinger, R., Garrido Rodriguez, G., & Kidd, E. (2025). Word order flexibility in Pitjantjatjara, across genres and generations. In C. O'Shannassy, J. Bednall, & D. Angelo (Eds.), *Projecting voices: Studies in language and linguistics in honour of Jane Simpson*. Asia-Pacific Linguistics.