The Role of Input in Language Revitalization:
The Case of Lexical Development

William O'Grady
University of Hawai'i at Mānoa

Raina Heaton
University of Oklahoma

Sharon Bulalang
University of Hawai'i at Mānoa

Jeanette King
University of Canterbury

Immersion programs have long been considered the gold standard for school-based language revitalization, but surprisingly little attention has been paid to the quantity and quality of the input that they provide to young language learners. Drawing on new data from three such programs (Kaqchikel, Western Subanon, and Māori), each with its own particular motivation, objectives, and pedagogical practices, we examine a key component of this revitalization strategy, namely the amount and type of lexical input that children receive. Our findings include previously unknown facts about the number of words that children in these programs hear per hour, the ratio of word tokens to word types, and the skewed frequency distribution of the particular words that make up the input. We discuss our findings with reference both to comparable measures for first language acquisition in a home setting and to their relevance for pedagogical strategies in the classroom.

1. Introduction[1] Language revitalization takes many forms, ranging from efforts to increase awareness of a community's linguistic traditions to attempts to ensure that an endangered language is transmitted to the next generation of young people. In the latter case, on which we focus here, a common strategy for language revi-

---

talization takes the form of a school-based immersion program. Because the aim of such programs is to promote language acquisition, they can benefit from planning that is informed by the ample research literature on language acquisition and language pedagogy.

We concentrate here on what is arguably the single most important factor for linguistic development: the availability of ample high-quality 'input' in the form of speech. Because the input consists of talk by fluent speakers, it is at least partly under external control, making it possible not only to assess it but also even to modify it in ways that could enhance the opportunity for successful acquisition of the language.

In the case of many immersion programs, the primary source of this input over the course of the school day typically comes primarily (if not exclusively) from the teacher's interaction with his or her class. Our goal here is to conduct a preliminary investigation of 'teacher talk' in three language revitalization programs, which we then compare to caregiver speech in a monolingual home setting. As will become clear as we proceed, we focus on two factors that are widely used for the assessment of the vocabulary to which learners are exposed: its *quantity* (as measured by the total number of words that learners may hear) and its *diversity* (as measured by the number of distinct words that they encounter).

We begin in the next section by briefly summarizing the relevance and importance of vocabulary studies to language development, as well as some of the major findings of research in this area. Section 3 describes the methodology that we employed in our study. Section 4 reports our results, followed by a discussion in section 5 and some general concluding remarks in section 6.

2. Vocabulary development　The role of input in lexical development in first-language acquisition has been studied from two related perspectives. The first and older line of research concentrates on quantity (the amount of speech to which learners are exposed), whereas the second approach adds a focus on diversity (roughly, the number of different words that are encountered). We will briefly consider each in turn. Although this literature naturally focuses on preschool children, the lexical input relevant to the first years of a school-based program can be expected to include many of the same person-denoting, thing-denoting, and action-denoting words (O'Grady 2005: 41ff) that are needed for everyday communication in a home setting. The lexical needs of preschoolers learning their first language and young school-age children who are just beginning to learn a second language can therefore be assumed to overlap to a fairly high degree.

2.1 Quantity　An important milestone in the study of vocabulary development was a groundbreaking research project undertaken by Hart & Risley (1995; 1999), who made monthly one-hour recordings of forty-two children growing up in monolingual English-speaking families in the United States. The recording sessions began when the children were seven to nine months old and continued for two and a half years.

Drawing on extrapolations from the monthly samples, Hart & Risley reported vast differences in the amount of language to which individual children were

exposed. At one extreme were children from more talkative families, who heard around 30,000 words per day. In contrast, children from the least talkative families heard far less speech – around 8,600 words per day on average, according to Hart & Risley's estimates (1995: 132).[2]

A subsequent and even more ambitious study by Gilkerson et al. (2017) gathered day-long (twelve-hour) samples once a month in a total of 329 families over a period of six to thirty-eight months. (At the time of the recordings, the children ranged in age from two to thirty-eight months.) The resulting 38,556 hours of data, analyzed using LENA technology,[3] corroborated the essentials of Hart & Risley's work. There were significant differences across families in the amount of input to which children had access, although the overall word counts in Gilkerson et al.'s study were somewhat lower than those of Hart & Risley, ranging from around 11,000 to 15,000 per day on average. (Nonetheless, a small number of parents in their study did produce approximately 20,000 words a day.)

The principal interest of these studies lies not so much in the word counts themselves as in the impact of input on vocabulary learning. At the age of thirty months, children from the most talkative families in Hart & Risley's study had vocabularies more than twice the size of the vocabulary of children from the least talkative families (1995: 164). Moreover, in the subsequent six months, the children from the highly talkative families went on to learn more than twice as many new words as their peers did. Similar correlations have been reported by Hoff (2003) and Fernald et al. (2013).[4]

Languages other than English   The relevance of input to lexical development has also been documented for languages other than English. Weisleder & Fernald (2013) investigated language acquisition in a group of twenty-nine Spanish-speaking Latino children in the United States (from families with the same socioeconomic status). Their results revealed "striking variability" in the amount of adult speech addressed to the children in samples collected when they were nineteen months old. Some chil-

---

[2] Hart & Risley's work has generated controversy over the relationship between a family's socioeconomic status and its linguistic practices, as well as the possible consequences of this relationship for children's later academic achievement (see, e.g., Kuchirko 2019). We use Hart & Risley's findings, and those of other scholars who have conducted similar studies, only to support the claims (a) that there are differences among children in terms of the amount of language they hear and (b) that – not surprisingly – those differences are correlated with children's vocabulary size. Put simply, the more words learners hear, the greater the opportunity to increase the size of their vocabulary. For further discussion, see Golinkoff et al. (2019).

[3] LENA (Language ENvironment Analysis) consists of software that is able to automatically generate reliable estimates of adult words counts, child vocalization frequency, and conversational turn-taking (Gilkerson et al. 2008; Xu et al. 2009; Ganek & Eriks-Brophy 2018). Although LENA's categorizations are highly accurate and yield word counts for adult speech that are very close to those of human transcribers (Gilkerson et al. 2017: 251), we cannot discount the possibility that the difference in methodologies might be responsible for a small percentage of the divergence in the estimates reported in the two studies.

[4] The consequences extend beyond the lexicon: Vocabulary size also predicts syntactic development as well as various types of cognitive development (e.g., Montag et al. 2018: 399, 402 and the references cited there).

dren heard as many as 29,000 words in the course of a day, and some fewer than 2,000. Crucially, the children to whom more speech had been directed had substantially larger vocabularies six months later and were quicker to recognize words. A similar finding has been reported by Shneidman & Goldin-Meadow (2012), who examined the relationship between input and development in a Yucatec Maya community in Mexico. Based on a study of fifteen families, they reported that the amount of speech directed to children at the age of twenty-four months was strongly correlated to the size of their vocabularies eleven months later.

Bilingual settings   Almost by definition, immersion settings involve a commitment to bilingual development, for which the role of input is of vital importance given the need to acquire the vocabulary of two distinct languages. Not surprisingly, the literature on this subject confirms that lexical development in a bilingual setting is closely tied to the quantity and quality of the input to which children are exposed. A study along these lines was conducted by Hoff et al. (2012), who investigated the development of Spanish and English in forty-seven children at three points in their development (ages 1;10, 2;1, and 2;6).

Hoff et al. uncovered a strong effect of dual language input on vocabulary size. Children who had more exposure to Spanish attained higher Spanish vocabulary scores than children with balanced exposure to the two languages, who in turn had higher scores than children who had been predominantly exposed to English. In contrast, the latter group of bilinguals did better on English vocabulary tests than the children who had predominant exposure to Spanish or balanced exposure to both languages. As Hoff et al. note, "the proportion of home language input in English was positively and significantly correlated to every measure of English development at every time point and negatively related to every measure of Spanish at every time point" (2012: 19).

2.2 Diversity   Early vocabulary studies focused on the number of words to which children are exposed. However, more recent work has identified a potentially more important variable for predicting vocabulary growth, namely lexical diversity – the number of different words that language learners hear.

One indication of the importance of this factor comes from Pan et al.'s 2005 study of the correlation between maternal speech and vocabulary production by children aged one and three in 108 families. The authors report that the number of different words used by mothers was the best predictor of child vocabulary growth (Pan et al. 2005: 776). Although maternal talkativeness was positively correlated with the number of different words that are used (the more mothers talked, the more different words they produced), it did not have an independent effect on growth in vocabulary production. This finding confirmed earlier work along the same lines by Weizman & Snow (2001), which also showed that diversity of word use is a better predictor of child vocabulary outcome than the mere amount of input per se.

A further influential study was conducted by Rowe (2012), who measured the long-term effect of parental language use in fifty families by gathering speech samples when the children were eighteen, thirty, and forty-two months old. She found that

whereas the sheer quantity of words that a child hears is important in the second year of life, diversity and sophistication of vocabulary become the better predictors of lexical growth in the third and fourth years. Jones & Rowland (2017) confirmed this result using a computer model based on input samples from sixteen different mothers. Hoff (2006) and Jones & Rowland (2017: 2) provide reviews of the literature on this subject.

These and other studies have produced a sizeable amount of information about the role of input in language learning, creating an opportunity that has not previously been exploited to better understand the advantages and challenges of school-based revitalization programs. The primary point of interest lies in the importance of caregiver input to vocabulary growth, a key factor in language acquisition and, therefore, in language revitalization as well. Our study has two specific goals:

    i.   to compare the input available to children in different types of immersion programs with the input in first-language contexts.

    ii.   to analyze the input available in the immersion programs with a view to better understanding its relevance to language planning, curriculum design, and pedagogical practice.

We describe our study in the next section.

3. The immersion study   Our study focused on three immersion programs that share characteristics important to our research objectives: (1) all have the expressed goal of creating and/or maintaining proficiency in an endangered language, (2) all are school-based, and (3) the students in the programs are preadolescent children.

3.1 Participating programs   Two of the programs, a Kaqchikel (Maya) school in Guatemala and a Māori school in New Zealand, offer a classic immersion curriculum designed for children who initially have limited or no proficiency in the endangered language.

Kaqchikel (Mayan; ISO 639-3: cak)
Institution and type of immersion program:
    Nimaläj Kaqchikel Amaq' in Chimaltenango, Guatemala – a partial immersion program. Children receive instruction through the medium of Kaqchikel in math, art, computer use, physical education, and Kaqchikel language arts, for a total of approximately two hours per day, and are encouraged to use the language during recess and lunch time. All other instruction and activities are in Spanish.
Brief sketch of the language:
    Kaqchikel is a verb-initial language belonging to the K'ichean branch of the Mayan language family. It is known for its complex verbal morphology,

which can be used to express entire sentence-like meanings. Its phonology includes unusual uvular and glottal consonants, as well as contrasts involving glottalization.

Vitality level:

Vulnerable, per the Catalog of Endangered Languages, based on data from the *Atlas Lingüístico de Guatemala* (Richards 2003). About half the ethnic population speaks the language, but a shift to Spanish monolingualism is particularly pronounced in urban communities like the one in which our school was located, where Kaqchikel has not been widely used in public for at least two generations (Heaton & Xoyón 2016). Language shift is also evident in large rural Kaqchikel towns like San Juan Comalapa, where it is increasingly the case that younger generations do not speak the language fluently. However, the language continues to be the primary means of communication in many of the smaller rural communities (*aldeas*).

Dominant language in the region:

Spanish. There are few opportunities to use Kaqchikel in Chimaltenango outside of school.

Cohorts participating in the study:

First through fourth grade (math class only), totaling twenty-six students and averaging thirteen students per class. The children were all ethnically Kaqchikel but not from Kaqchikel-speaking homes. Nearly all the students come from severely impoverished situations.

Teacher background:

We recorded a single teacher in his early twenties from San Juan Comalapa, who is a native speaker of Kaqchikel as well as Spanish. He taught math in Kaqchikel at all grade levels.

Māori (Polynesian; ISO 639-3: mri)

Institution and type of program:

A total immersion school in Christchurch, South Island, New Zealand; all instruction took place in Māori.

Brief sketch of the language:

Māori is a lightly inflected verb-initial language, with a phonology consisting of ten consonants and five vowels (for which there is a length contrast) and a CV syllable template. It belongs to the Malayo-Polynesian branch of the Austronesian family.

Vitality level:

Endangered, per the Catalog of Endangered Languages, based on data from the *Oxford Handbook of Endangered Languages* (King 2018). About 21% of the Māori population is able to converse in the language, with roughly 40% of those over sixty-five claiming fluency. Domains of use include traditional temples (*marae*), churches, language nests, immersion schools, radio, and television.

Dominant language in the region:

> English

Cohorts participating in the study:

> Twenty-five year-five students aged nine and ten. Most of the children had been in Māori immersion schooling since the age of five. About half received semiregular exposure to Māori in the home or community. The school hosting the program was classified as 'Decile 3,' which places it in the 30% of schools with a high proportion of students from low socioeconomic communities.

Teacher background:

> The teacher we recorded is a second-language speaker of Māori in her mid-thirties, who at the time was in the final semester of her first year of teaching. Her first language is English.

The third program differed from the first two in being designed to maintain a language that is already spoken by its students but is being used by an increasingly small number of families in its traditional territory.

Western Subanon (Philippine; ISO 639-3: suc)

Institution and type of program:

> Malayal Community School in the province of Zamboanga del Norte on Mindanao is a total immersion program. It uses Western Subanon as the medium of instruction for language arts, history, mathematics, music, art, physical education, and values education, as well as for training in English and Tagalog.

Brief sketch of the language:

> Western Subanon is a verb-initial language, belonging to the Greater Central Philippine branch of the Austronesian family. It uses a complex system of prefixing, infixing, and suffixing to express aspect, modality, and a four-way contrast in voice. Its phonological inventory contains fifteen consonants and five vowels; most syllables have a (C)V(C) template.

Vitality level:

> Endangered, per the criteria of the Catalogue of Endangered Languages as applied by a coauthor of this paper who is a native speaker of the language and a trained linguist. It is no longer the dominant language in most areas where it was once spoken, and it is not used by parents when speaking to their children in those areas. However, the particular region in which we conducted our study is exceptional in that Western Subanon is widely used and parents still speak it to their children.

Dominant languages in the region:

> Cebuano, Chavacano, Tausug

Cohorts participating in the study:

> Thirty-five first-grade students and thirty-six second-grade students. All the children are native speakers of Western Subanon whose parents speak to them in the language. However, the children and their parents are also flu-

ent to varying degrees in Cebuano and Chavacano. Students come from impoverished situations.

Teacher backgrounds:

Both the first-grade and second-grade teachers are native speakers of Western Subanon and are native to Malayal. Both teach all subject areas in their respective classes.

In addition, for purposes of comparison, we have made extensive use of data on child-directed speech in monolingual English-speaking families, for which we relied on studies available in the published literature. We do not believe that immersion programs can be expected to replicate the conditions under which family-based first language acquisition takes place. Nonetheless, acquisition of a first language in a monolingual family provides a useful baseline in its own right since it represents the one setting in which language acquisition is invariably successful. As we will see, comparisons with this setting not only prove to be helpful but at times also yield pleasantly surprising results.

The data for Kaqchikel, Māori, and Western Subanon were collected between 2016 and 2017, and therefore may no longer be representative of the current situation as changes in curriculum, policies, and personnel inevitably take place over the course of time. Moreover, we wish to stress that our goal is not to evaluate or pass judgement on any of the programs, but only to contribute to a better understanding of the workings of school-based language revitalization and to make available information that might be of use to other programs.

3.2 Methodology In order to gather speech samples, we arranged for teachers to wear a recording device (a Zoom H4n recorder with an external lavalier mic) during the course of the teaching day for the periods of time reported in Table 1 in Section 4 below. Because we were conducting a study on input from classroom caregivers only, no attempt was made to record the speech of the children in the programs, a project that was impractical for logistical reasons since it would have required an entirely different protocol and set of permissions. We recognize of course that in many language acquisition settings, children learn a great deal from interactions with other children. However, the question of whether this generalization applies in the case of immersion classrooms calls for independent verification given that, typically, few of the children are initially fluent in the target language (The Western Subanon program is an exception in this regard).

Once the recordings had been made, they were transcribed by fluent native speakers according to the standard orthographic conventions for the language in question. Transcribers were trained to segment the audio at the utterance level, treating pauses and conversational turns as boundaries. The transcripts were also tagged for 'content words' (nouns and verbs) as well as for information relevant to other planned studies. All tagging was done in time-aligned dependent tiers in ELAN, an annotation tool for audio and video recordings that is widely used for language documentation.

Basic measurements, such as total number of words and total duration of the recordings, were taken directly from the internal metrics of ELAN. Other calculations, such as the total number of unique words, were generated using ELAN's search and export functions and then compiled by the authors.

For the purposes of our counts, we took a word to be a form consisting of a root and any associated affixes. This definition works well for the languages in our study and for our current purpose, which is to calculate the quantity and diversity of *vocabulary items*.[5] A follow-up paper, currently in preparation, will report on the occurrence and distribution of morphosyntactic features with a view to assessing the extent to which the input is rich enough to support grammatical development.

4. Findings Our analysis of the raw data focused on four criteria: the number of hours of actual teacher talk, the mean number of words per hour of contact time, the degree of lexical diversity, and the frequency distribution of the words to which the children are exposed.

4.1 Speech time We requested that each program carry out audio recording for a period of two consecutive school weeks. However, for practical reasons relating to differences among the individual programs, their teaching schedules, and the length of their school day, the amount of recording varied somewhat from school to school. For example, the Western Subanon school provided one week of audio recording from each of two classrooms. In addition, because the Kaqchikel school involved a partial immersion program, the number of hours of recording was naturally smaller than for the two full-immersion programs. Table 1 reports on the total period of time during which the teachers were available to interact with their students in the language (henceforth 'school-day contact time') and on the number of hours of actual language use by the teachers during that period.

---

[5] The languages differ in terms of how they build their words – especially their verbs. Verbs in Māori are typically uninflected but are accompanied by particles (written as separate words) that provide information about tense, aspect, and modality. In contrast, verbs in Western Subanon require voice marking and may also be inflected for number agreement, aspect, and modality. The Kaqchikel verb is even more morphologically complex; it is always inflected for agreement with its subject and direct object as well as for tense/aspect/modality, and it may also carry marking to indicate class membership and derivational processes of various sorts. However, these differences appear to be reflected in the number of inflectional affixes rather than in the number or diversity of content words (nouns and verbs), which are the foundation of children's lexicons.

Table 1. School-day contact time versus speaking time

Language	School-day contact time	Speaking time	Percentage
Kaqchikel	22 hrs, 30 mins	3 hrs, 5 mins	13.7%
Māori	51 hrs, 20 mins	10 hrs, 44 mins	20.9%
W. Subanon-1	15 hrs	3 hrs, 50 mins	25.6%
W. Subanon-2	20 hrs	4 hrs, 22 mins	21.8%

These ratios and the variation that they represent may seem surprising, but they are apparently not out of line with the 'time-to-talk' ratios found in naturalistic situations. For instance, Van de Weijer (2002) reported that in his ninety-one-day study of a German- and Dutch-speaking household, the mean daily recording time of 7.9 hours included an average of just two hours and thirty-three minutes of actual speech (32.28%). Moreover, based on a study of 396 English-speaking university students, Mehl et al. (2007) estimated that the average number of words spoken in a seventeen-hour day is about 16,000. Assuming a speech rate of 150 words per minute, in accordance with estimates by the National Center for Voice and Speech, this comes out to around 106 minutes of speech per day on average (9.62% of a waking day). If anything, then, the amount of speech heard in an immersion classroom on a per-hour basis may well be greater than what would be encountered in a noneducational context.

4.2 Number of words   Given the relatively small portion of the school day during which children are directly exposed to their teacher's speech (here 13.7%–25.6% of their school-day contact time), the nature and quantity of what children do hear become extraordinarily important. Table 2 summarizes the total number of words in the speech samples that we collected.[6]

Table 2. Number of words in the recordings of teacher speech

Language	Words
Kaqchikel	21,193
Māori	88,160
W. Subanon-1	24,407
W. Subanon-2	32,181

---

[6] These word counts exclude partial words, unintelligible speech, words in other languages, and proper nouns.

By dividing the totals in Table 2 by the school-day contact time, we were able to arrive at the per-hour estimates of speech input reported in Table 3.

Table 3. Number of words per hour of school-day contact time

Language	School-day contact time	Words per contact hour
Kaqchikel	22.5 hrs	942
Māori	51.33 hrs	1,718
W. Subanon-1	15 hrs	1,627
W. Subanon-2	20 hrs	1,609

Our teachers provided an average of 942–1,718 words per hour of talk during the recording period. As illustrated in Table 4, these figures fall within the range reported for caregiver speech in first-language home settings, which we know is ideal for successful language acquisition.

Table 4. Number of words per hour of contact time in English first-language home settings

Study	Words per hour of contact time
Hart & Risley (1995: 132)	620-2,150
Brent & Siskind (2001)	2,348.7[a]
Gilkerson et al. (2017: 259)	1,025
Hoff & Naigles (2002: 426)	2,688.6
Rowe (2012: 1767)	2,375
Roy et al. (2009: 3)	2,401.6[b]
Shneidman et al. (2013: 678)	2,404[c], 2,063[d]

[a] As reported by Quick et al. (2019: 123)

[b] Including words spoken by the child himself

[c] Single-speaker households

[d] Multiple-speaker households

The per-hour input for Kaqchikel falls at the lower end of the range reported for English, most closely approximating the mean number of words produced by the caregivers in Gilkerson et al.'s 2017 study. The input counts for Māori and Western Subanon are noticeably higher and are comparable to the mid-range of those that have been reported for English. Thus, despite the differences in age and setting, these results suggest that per-hour input in a school context can be similar *in quantity* to what is available on an hour-by-hour basis in a first-language home setting.

4.3 Lexical diversity We turn next to the matter of lexical diversity, as reflected in the number of different words (also known as 'unique words' and 'word types') to which children are exposed in the immersion programs that we studied. Table 5 offers a preliminary estimate of lexical diversity in the four classrooms that we sampled.[7]

Table 5. Number of word tokens and word types

Language	Tokens	Types	Token-to-type ratio
Kaqchikel	21,193	1,214	17 to 1
Māori	27,263	1,041	26 to 1
W. Subanon-1	24,407	1,652	15 to 1
W. Subanon-2	32,181	2,685	12 to 1

Care must be taken in calculating and comparing token-to-type ratios as these are heavily affected by sample size. If, for example, the sample consisted entirely of the preceding sentence in this paragraph, the token-to-type ratio would be 1 to 1 since each word is distinct from all the other words. Obviously, this ratio would change dramatically if it was calculated for the entire paper.

The following graph, modified from Montag et al. (2015: 4), shows the mean number of unique words (types) as a function of the total number of words (tokens) in samples of child-directed speech from the CHILDES English database.[8]

---

[7] In order to maintain comparability in terms of corpus size (given that larger corpora result in larger token-to-type ratios), we used the smaller Māori sample that had been obtained for the first stage in our research on this language, which consisted of 27,263 words. For our full 88,160-word Māori corpus, the token-to-type ratio was 54:1.

[8] Calculations of this type depend on a variety of factors, including the treatment of pluri-functional words. In English, *by* can indicate location (*sit by the tree*), time (*be here by 3:30*), agency (*chased by a dog*), means (*go by bus*), and so on; in Māori, *ki* can be used to mark direction, location, an instrument, or a direct object. If items like these are counted as a single word (as we consistently chose to do), the total number of distinct lexical items will be lower than if each usage was taken to involve a separate word.
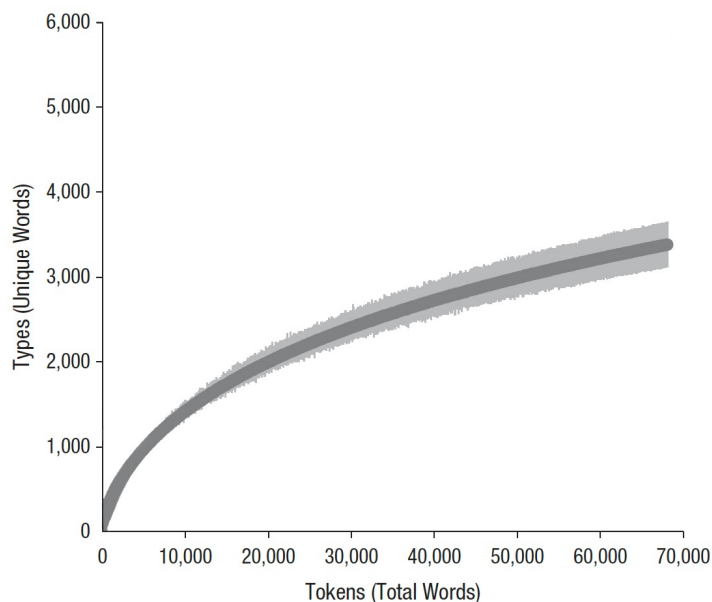
Figure 1. Token-to-type ratios for child-directed speech in the CHILDES database

Based on a 20,000-word English corpus of child-directed speech (about the size of the corpora for the three languages in our study), Montag et al. (2018: 378) estimated a token-to-type ratio of 8.78 to 1. This ratio points to a substantially higher degree of lexical diversity than we found in our comparably sized samples of teacher talk for Māori (26 to 1), Kaqchikel (17 to 1), or Western Subanon (15 to 1 and 12 to 1).[9]

4.4 The distribution of words   It has long been known that the use of words in a language's vocabulary complies with Zipf's Law.

> **Zipf's Law** (paraphrased; see Zipf 1949)
> The words used in natural speech are heavily skewed with respect to their frequency.

What this means is that the second most frequent word in a corpus may well be used just half as often as the most frequent word, the third most frequent word may be used just a third as often, the fourth a quarter as often, and so on. The result is a

---

[9] In the case of Kaqchikel, lexical diversity may have been affected by the fact that the classes that we recorded focused on math.

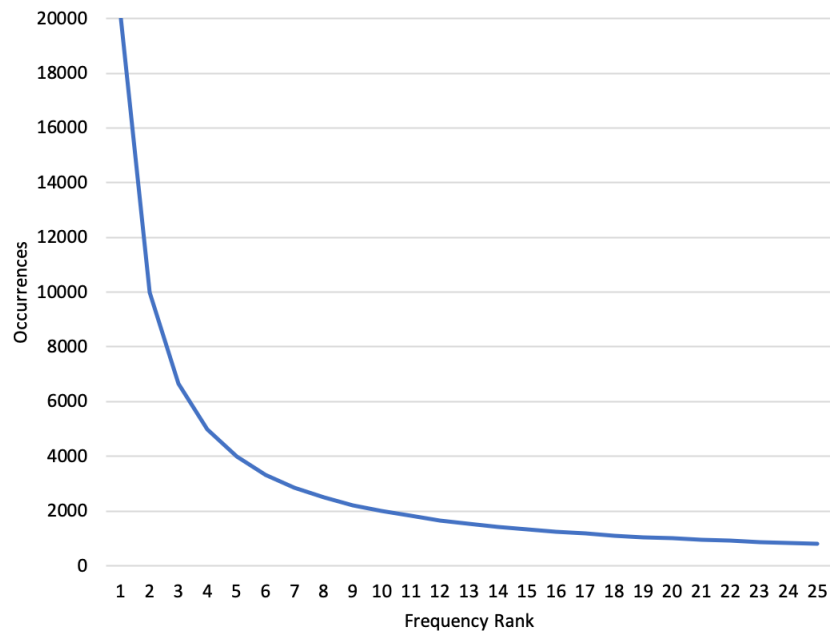trajectory of lexical usage, which – in its idealized form – looks something like the curve in Figure 2.

Figure 2. Idealized Zipfian curve

The key point of importance for language learning is that "relatively few words are used frequently […] while most words occur rarely, with many occurring only once in even large samples of texts, falling on the long tail" of the curve (Yang 2016: 18). This cannot but affect the opportunities for learners to extend their vocabulary. The fact that a very large proportion of the words in a language are encountered very infrequently means that learners will have only fleeting exposure to most lexical items.

The curve representing the relative frequency of the hundred most commonly used words in each of our corpora is given as a percentage of the total word count in Figure 3.
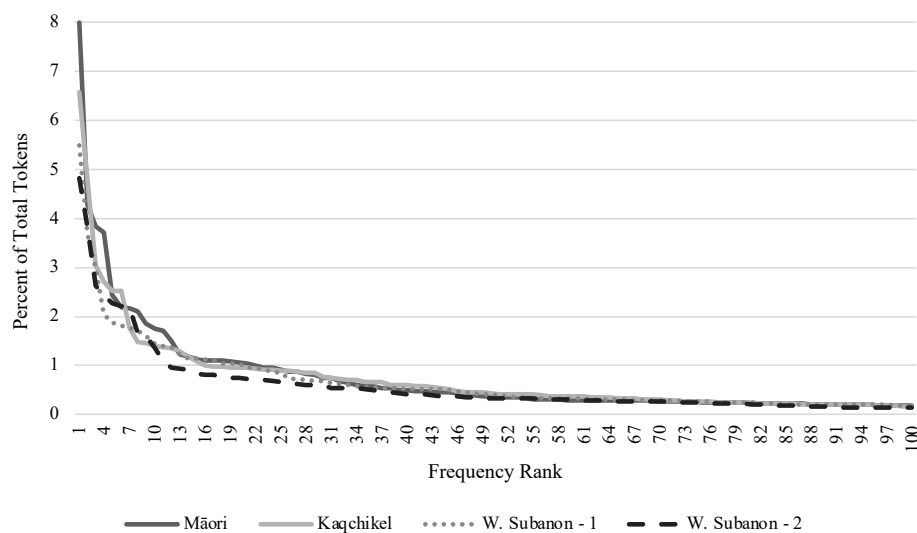
Figure 3. The relative frequency of the 100 most commonly used words in our Māori, Kaqchikel, and Western Subanon corpora

All three corpora exhibit a similar frequency distribution, with the rate of usage quickly falling to less than one percent of the total corpus for all but the most frequent twenty words in each language. To see how this hundred-word sample compares to the rest of the corpus, consider the percentages given in Table 6.

Table 6. Percentage of the total corpus represented by the most frequent 25, 50, and 100 word types in each sample

	25 most frequent words		50 most frequent words		100 most frequent words	
	% corpus	% types	% corpus	% types	% corpus	% types
Kaqchikel	44%	2.1%	60%	4.1%	74%	8.2%
Māori	49%	1.5%	64%	3.1%	77%	6.2%
W. Subanon -1	41%	1.5%	55%	3%	68%	6.1%
W. Subanon -2	37%	0.9%	49%	1.9%	61%	3.7%
English[a]	35%	0.5%	50%	1%	65%	1.9%

[a] These figures come from Quick et al.'s 2019 study of child-directed speech in the 257,480-word Brent corpus (Brent & Siskind 2001).

For all five corpora across four languages (including English), the fifty most frequent words make up roughly 50%–60% of the corpus. This means that approximately half of what teachers said to their students in the sample period consisted of the same fifty words. Moreover, those fifty words represent only between 1% and 4% of the unique words (word types) in each corpus. Taken together, these facts – which are remarkably similar across the languages – present obvious challenges for vocabulary development, especially in contexts where there is limited exposure to the language to begin with.

It is also worth noting that the most frequently occurring items in each language tend to be closed-class words (numerals, determiners, prepositions, and the like). This is the case for eight of the ten most common words in Māori and for nine of the ten most common words in Kaqchikel, in Western Subanon, and in the Brent corpus for English (Quick et al. 2019: 128). Indeed, fewer than half of the hundred most frequent words in each corpus consist of "content words" (nouns or verbs), as shown in Table 7.

Table 7. Number of content words in the most frequent 100 words

Language	Words
Kaqchikel	47
Māori	27
W. Subanon-1	29
W. Subanon-2	30
English[a]	33

[a] Brent & Siskind 2001

Here again there are evident implications for vocabulary development: the fact that a large proportion of the most frequently heard words in a language are function words rather than content words may facilitate certain aspects of morphosyntactic development (a matter to which we will turn in a future report), but it does so at the expense of lexical diversity – with potential consequences for both speech and comprehension.

Finally, we calculated the number of word types that appeared fewer than three times in each of the three corpora from our participating language revitalization programs.

Table 8. Number and percentage of word types with fewer than three tokens

Language	Tokens	Types
Kaqchikel	682	56%
Māori	638	39%
W. Subanon-1	927	56%
W. Subanon-2	1,425	53%

As illustrated here, the proportion of infrequent words was similar across the corpora, with ~40%–60% of all unique words appearing fewer than three times in the speech recorded for each language.

All of these findings are reflections of the skewed distribution of lexical items that is typical of language. This should not be surprising; Zipf's Law is universal (e.g., Sorell 2012; Yang 2016: 18), and we expect to see its effects in the distribution of words in all languages if sample sizes are sufficiently large. We will consider the consequences of our findings in the next section.

5. Discussion and implications The data discussed in the preceding sections reveal a number of broad quantitative similarities in the type of input found in immersion classrooms and the type of input to which English-speaking children have access in a monolingual home setting. These similarities include:

- high overall time-to-talk ratios
- a per-hour word total ranging from ~900 to ~1,700, compared to ~600 to ~2,700 for a monolingual English home setting
- a type-to-token ratio ranging from 26:1 to 12:1 for speech samples in the 20,000–35,000 word range compared to 8:1 for a monolingual English home setting
- a Zipfian effect, with the result that the fifty most frequent lexical items make up about half of all words encountered in the input (all settings).

However, it is important not to confuse similarity with parity. Children in the Kaqchikel immersion program may well have heard an average of ~1,000 words per hour, but that does not change the fact they were spending just two to three hours per day (ten to thirteen hours per week) in an environment where there was a chance to hear the language spoken, compared to twelve to fourteen hours per day (eighty-four to ninety-eight hours per week) for a child in a typical monolingual setting.

This contrast becomes particularly impactful when we take into account the effects of Zipf's Law, which guarantees a highly skewed distribution of vocabulary items – one consequence of which is that many content words occur very infrequently. This fact takes on special importance in the context of vocabulary learning. As Montag et al. (2018: 383) put it, "the specific words at the head of the distribution

are very frequent, but most of the words that children need to learn—the long tail of the distribution—are infrequent."

Early research on children's lexical development often reflected a fascination with the phenomenon of 'fast mapping,' which allows a new word to be acquired upon a single exposure (Carey & Bartlett 1978; Markson & Bloom 1997). It is now widely agreed that this scenario does not represent the complete picture and that multiple exposures, possibly over a period of months and even years, are often required for full acquisition of new words (Carey 2010: 4). As Harris et al. (2011: 51, 57) note, first exposure to a word yields no more than "a cursory understanding of word meaning; repeated exposures to a new word in varied contexts, or the provision of definitions to which children can relate, lead to a deeper, more nuanced understanding of word meaning."

Various studies have confirmed these observations. Schwartz & Terrell (1983) found that children aged twelve to eighteen require, on average, ten to twelve exposures to a novel word to be able to produce it appropriately. In a study of 120 Dutch-speaking children aged five to ten, Ameel et al. (2008) found that knowledge of words for familiar household items (e.g., glass, cup, bottle, box, container, tube) continued to develop well into adolescence. Neuman & Wright (2014: 10) report that 24 exposures to a novel word were required for it to be successfully remembered by 80% of the sixty 4-year-olds in their study. Childers & Tomasello (2006) suggest that eight exposures might suffice, based on a study of thirty-six 2-year-olds, but noted that exposure to novel words on different days is more important than just the number of exposures (see also Childers & Tomasello 2002).

Similar findings have been reported for second language learners. In a review of the literature on this topic, Nation (2014: 2–3) reports that experimental work has yielded estimates varying from between "two or three" and ten exposures in order to ensure learning of a new vocabulary item. Nation himself suggests twelve as a "moderately safe" minimum goal.

On average, children acquiring English in a first-language context have vocabularies of around 6,000–10,000 unique words (i.e., word types rather than tokens) by the time they are six years old (Bloom 2001: R5; Medina et al. 2011: 9014; Segbers & Schroeder 2016: 298–299). For the most part, the words that have been acquired at this point in their lives are those to which they have been exposed most frequently (Goodman et al. 2008: 524).

How much input would a child have to receive in order to acquire 6,000-10,000 words during the first few years of life? And how much would be required to learn the many hundreds of additional words per year that is characteristic of lexical development (Bloom & Markson 1998; Segbers & Schroeder 2016)? It is impossible to give a precise number, of course, but a calculation done by Nation (2014) is worth mentioning. As part of a far-reaching analysis of the type of input available to learners of English, he found that just 6,457 of the 9,000 most common words in English occurred in a two-million-word sample of spoken English in the British National Corpus. Thus, even exposure to two million words of input does not suffice to ensure that a learner will encounter more than two-thirds of the 9,000 most frequent words

in his or her language. (And, of course, many of those words may be encountered too infrequently to actually be learned.)

Yet another issue that must be taken into account involves the recent finding that participation in one-on-one conversations is a major predictor of vocabulary size (e.g., Romeo et al. 2018). Although our methodology was not designed (or able) to identify to whom the teachers were talking when they spoke, we know from anecdotal accounts that the teacher in one of the programs tended to interact with students on an individual basis rather than as a group. It seems reasonable to assume that the interactional style of individual teachers will vary across all types of class-rooms, not just those devoted to immersion programs. However, this variable seems to be especially important in the latter case since the benefits of one-to-one conversa-tions have to be weighed against the effect on the overall amount of language that is available to the entire class. This matter calls for careful investigation.

6. Concluding remarks   Even in the most favorable of circumstances, a typical immersion program is probably in session no more than twenty-five hours of a ninety-eight-hour week – about 25% of a child's total waking hours. Interestingly, this estimate aligns rather closely with the minimum ratio of exposure that is often recommended for successful acquisition of a second language, which falls in the 25%–30% range (Genesee 2007; Baker 2014: 38). This leaves a rather small margin of comfort for school-based immersion programs, especially when the likelihood of dramatically reduced exposure to the target language during weekends and vacation periods is taken into account.

It is clear that the success of immersion programs lies in finding ways to increase the types of exposure and interactions that are available to young language learners, particularly in contexts where the language is not widely spoken in the home and community. The key to this effort, we believe, lies in informed planning, careful im-plementation, and regular assessment. Four considerations require special attention.

First, it is important to have at hand a basic lexicon of the language that identi-fies the vocabulary items that are most essential to the type of setting in which the language is likely to be used. Many words of this type, such as the names for body parts, common objects in the environment, and basic actions, have parallels across languages, offering a possible common starting point. A potentially useful resource in this regard is Wordbank (http://wordbank.stanford.edu/), an open database with information about early vocabulary gathered from more that 75,000 children repre-senting twenty-nine different languages.

Second, it is essential that immersion programs track children's lexical develop-ment. Indeed, we venture to say that there is no immersion program anywhere in the world that would not benefit from information about the nature and extent of its students' vocabulary knowledge. There are many instruments for conducting this sort of assessment, the most popular of which is a picture-naming task (see Hoffman et al. 2014 for a critical review).

Heaton & Xoyón (2016) report on such an assessment at the same Kaqchikel immersion school at which we conducted our research (although with an earlier co-hort of students). They tested thirty-seven students ranging in age from five to ten on

a picture-naming task consisting of forty-five items, most of which were thought to involve familiar words, including some displayed on the classroom walls. Mean success rates ranged from 34.2% in the preprimary class to just 52.79% for the second graders, despite their long-term exposure to those words.

A third essential point involves the need to incorporate lexical items into the curriculum in a way that maximizes the chances of acquisition. This will involve overcoming two challenges: (1) the relatively limited amount of contact time associated with school-based immersion programs and (2) the effects of Zipf's Law, which include a sharply descending rate of usage for all but the most frequent words. A very deliberate effort must be made to "flatten the curve" by ensuring that each word in the target set occurs at a rate that will provide enough exposure for learning and retention.

There is a large literature on this topic that suggests that various types of interventions can be effective. For example, a survey by Marulis & Neuman (2010) found that explicit strategies for teaching vocabulary coupled with multiple exposures in varied contexts were most effective, particularly for pre-K and kindergarten children (p. 318). It is important to note, however, that the existing literature is heavily focused on English reading comprehension.

The reality of language revitalization is quite different, since the primary goal is typically oral proficiency, with writing as a secondary (but related) skill. As with many aspects of language revitalization, strategies must be tailored to each specific context. Sapién & Hirata-Edds (2019) have offered some suggestions for using primarily oral language corpora for language revitalization.

A fourth point calling for action involves finding ways to increase the amount of input that children receive over the course of a day. In many cases, this will involve more talk by the teacher or other fluent speakers who can be brought into the classroom. This suggestion goes against the grain of much recent work on teacher talk (e.g., Hattie 2012), including some studies on teacher talk in second-language classrooms (e.g., Lubin n.d.). However, recommendations that call for less teacher talk typically do not take into account the special needs of a school-based revitalization program, in which the prospects of the language's survival may well depend almost entirely on classroom-based input.

An important strategy for addressing this challenge could well involve the design and use of literacy materials, which are known to play a major role in vocabulary development (e.g., Wasik et al. 2016). One obvious advantage of written materials is that they offer the opportunity to pre-plan the choice of words and the particular contexts in which they are used, as well as to control their distribution and frequency – factors that are virtually impossible to monitor and manage in the case of spontaneous speech.

In sum, educators need to be aware of both what they are doing with language and what the students in the classroom are learning from it. In the case of children acquiring a first language in a monolingual setting, essentially everything can be left to chance. Over the course of time, given a reasonable amount of exposure, those children will acquire their language, including its vocabulary, to a satisfactory level of proficiency. In the case of immersion programs for endangered languages,

in contrast, *nothing* can be left to chance. The stakes are too high, and the risks are too great to do anything other than engage in a careful program of teacher training, curriculum planning, and proficiency assessment. Vocabulary learning offers an ideal opportunity to put this policy into practice.

References

Ameel, Eef & Malt, Barbara & Storms, Gert. 2008. Object naming and later lexical development: From baby bottle to beer bottle. *Journal of Memory and Language* 58. 262–285. https://doi.org/10.1016/j.jml.2007.01.006

Baker, Colin. 2014. *A parents' and teachers' guide to bilingualism*. 4th edn. Clarendon, UK: Multilingual Matters. https://doi.org/10.21832/9781783091614

Bloom, Paul. 2001. Word learning. *Current Biology* 11. R5–R6. https://doi.org/10.1016/S0960-9822(00)00032-4

Bloom, Paul & Markson, Lori. 1998. Capacities underlying word learning. *Trends in Cognitive Sciences* 2. 67–73. https://doi.org/10.1016/S1364-6613(98)01121-8

Brent, Michael & Siskind, Jeffrey. 2001. The role of exposure to isolated words in early vocabulary development. *Cognition* 81. B33–B44. https://doi.org/10.1016/S0010-0277(01)00122-6

Carey, Susan. 2010. Beyond fast mapping. *Language Learning and Development* 6. 184–205. https://doi.org/10.1080/15475441.2010.484379

Carey, Susan & Bartlett, Elsa. 1978. Acquiring a single new word. *Papers and Reports on Child Language Development* 15. 17–29.

Childers, Jane & Tomasello, Michael. 2002. Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology* 38. 967–978. https://doi.org/10.1037/0012-1649.38.6.967

Childers, Jane & Tomasello, Michael. 2006. Are nouns easier to learn than verbs? Three experimental studies. In Hirsh-Pasek, Kathy & Golinkoff, Roberta (eds.), *Action meets word: How children learn verbs*, 311–335. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195170009.003.0013

Fernald, Ann & Marchman, Virginia & Weisleder, Adriana. 2013. SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science* 16. 234–248. https://doi.org/10.1111/desc.12019

Ganek, Hillary & Eriks-Brophy, Alice. 2018. Language ENvironment analysis (LENA) system investigation of day long recordings in children: A literature review. *Journal of Communication Disorders* 72. 77–85. https://doi.org/10.1016/j.jcomdis.2017.12.005

Genesee, Fred. 2007. A short guide to raising children bilingually. *Multilingual Living Magazine* 2, 18–21.

Gilkerson, Jill & Coulter, Kimberly & Richards, Jeffrey. 2008. Transcriptional analysis of the LENA natural language corpus. (Technical Report No. LTR-06-2.) Boulder: LENA Foundation.

Gilkerson, Jill & Richards, Jeffrey & Warren, Steven & Montgomery, Judith & Greenwood, Charles & Oller, D. Kimbrough & Hansen, John & Terrence, Paul 2017. Mapping the early language environment using all-day recordings and automated analysis. *American Journal of Speech-Language Pathology* 26. 248–265. https://doi.org/10.1044/2016_AJSLP-15-0169

Golinkoff, Roberta & Hoff, Erika & Rowe, Meredith & Tamis-LeMonda, Catherine & Hirsh-Pasek, Kathy. 2019. Language matters: Denying the existence of the 30-million-word gap has serious consequences. *Child Development* 90, 985–992. https://doi.org/10.1111/cdev.13128

Goodman, Judith & Dale, Philip & Li, Ping. 2008. Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language* 35. 515–531. https://doi.org/10.1017/S0305000907008641

Harris, Justin & Golinkoff, Roberta & Hirsh-Pasek, Kathy. 2011. Lessons from the crib for the classroom: How children really learn vocabulary. In Neuman, Susan & Dickinson, David (eds.), *Handbook of early literacy research*, vol. 3, 49–65. New York: The Guilford Press.

Hart, Betty & Risley, Todd. 1995. *Meaningful differences in the everyday experience of young American children*. Baltimore: Paul H. Brookes.

Hart, Betty & Risley, Todd. 1999. *The social world of children learning to talk*. Baltimore: P.H. Brookes.

Hattie, John. 2012. *Visible learning for teachers*. New York: Routledge.

Heaton, Raina & Xoyón, Igor. 2016. Assessing language acquisition in the Kaqchikel program at Nimaläj Kaqchikel Amaq'. *Language Documentation & Conservation* 82. 317–352. (http://hdl.handle.net/10125/24716).

Hoff, Erika. 2003. The specificity of environmental influence: Socioeconomic status affects early development via maternal speech. *Child Development* 74. 1368–1378. https://doi.org/10.1111/1467-8624.00612

Hoff, Erika. 2006. How social contexts support and shape language development. *Developmental Review* 26. 55–88. https://doi.org/10.1016/j.dr.2005.11.002

Hoff, Erika & Core, Cynthia & Place, Silvia & Rumiche, Rosario & Señor, Melissa & Parra, Marisol. 2012. Dual language exposure and early bilingual development. *Journal of Child Language* 39. 1–27. https://doi.org/10.1017/S0305000910000759

Hoff, Erika & Naigles, Letitia. 2002. How children use input to acquire a lexicon. *Child Development* 73. 418–433. https://doi.org/10.1111/1467-8624.00415

Hoffman, Jessica & Teale, William & Paciga, Kathleen. 2014. Assessing vocabulary learning in early childhood. *Journal of Early Childhood Literacy* 14. 459–481. https://doi.org/10.1177/1468798413501184

Jones, Gary & Rowland, Caroline. 2017. Diversity not quantity in caregiver speech: Using computational modeling to isolate the effects of the quantity and diversity of the input on vocabulary growth. *Cognitive Psychology* 98. 1–21. https://doi.org/10.1016/j.cogpsych.2017.07.002

King, Jeanette. 2018. Māori: Revitalization of an endangered language. In Campbell, Lyle & Rehg, Kenneth (eds.), *The Oxford handbook of endangered languages*, 592–612. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190610029.013.28

Kuchirko, Yana. 2019. On differences and deficits: A critique of the theoretical and methodological underpinnings of the word gap. *Journal of Early Childhood Literacy* 19. 533–562. https://doi.org/10.1177/1468798417747029

Lubin, Matthew. n.d. Six ESL teaching techniques to cut TTT [teacher talking time] and get your students talking. *FluentU English Educator Blog*. (https://www.fluentu.com/blog/educator-english/esl-teaching-techniques/) (Accessed 2021-08-21.)

Markson, Lori & Bloom, Paul. 1997. Evidence against a dedicated system for word learning in children. *Nature* 385. 813–815. https://doi.org/10.1038/385813a0

Marulis, Loren & Neuman, Susan. 2010. The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research* 80. 300–335. https://doi.org/10.3102/0034654310377087

McCarty, Teresa. 2013. Literacy and language revitalization. In C. Chapelle (ed.), *The encyclopedia of applied linguistics*. Boston: Wiley-Blackwell.

Medina, Tamara & Snedeker, Jesse & Trueswell, John & Gleitman, Lila. 2011. How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences* 108. 9014–9019. https://doi.org/10.1073/pnas.1105040108

Mehl, Matthais & Vazire, Simine & Ramírez-Esparza, Nairán & Slatcher, Richard & Pennebaker, James. 2007. Are women really more talkative than men? *Science* 317. 82. https://doi.org/10.1126/science.1139940

Montag, Jessica & Jones, Michael & Smith, Linda. 2015. The words children hear: Picture books and the statistics for language learning. *Psychological Science* 19. 1–8. https://doi.org/10.1177/0956797615594361

Montag, Jessica & Jones, Michael & Smith, Linda. 2018. Quantity and diversity: Simulating early word learning environments. *Cognitive Science* 42. 375–412. https://doi.org/10.1111/cogs.12592

Nation, Paul. 2014. How much input do you need to learn the most frequent 9000 words? *Reading in a Foreign Language* 26. 1–16. (https://files.eric.ed.gov/fulltext/EJ1044345.pdf.)

Neuman, Susan & Wright, Tanya. 2014. The magic of words: Teaching vocabulary in the early classroom. *American Educator* 38. 4–13.

O'Grady, William. 2005. *How children learn language*. Cambridge: Cambridge University Press.

Pan, Barbara & Rowe, Meredith & Singer, Judith & Snow, Catherine. 2005. Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Development* 76. 763–782. https://doi.org/10.1111/j.1467-8624.2005.00876.x

Quick, Nancy & Erickson, Karen & Mccright, Jacob. 2019. The most frequently used words: Comparing child-directed speech and young children's speech to inform vocabulary selection for aided input. *Augmentative and Alternative Communication* 35. 120–131. https://doi.org/10.1080/07434618.2019.1576225

Richards, Michael. 2003. *Atlas lingüístico de Guatemala*. Guatemala City: Universidad Rafael Landívar.

Romeo, Rachel & Leonard, Julia A. & Robinson, Sydney T. & West, Martin R. & Mackey, Allyson P. & Rowe, Meredith L & Gabrieli, John D. E. 2018. Beyond the 30-million-word gap: Children's conversational exposure is associated with language-related brain function. *Psychological Science* 29(5). 700–710. https://doi.org/10.1177/0956797617742725

Rowe, Meredith. 2012. A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development* 83. 1762–1774. https://doi.org/10.1111/j.1467-8624.2012.01805.x

Roy, Brandon & Frank, Michael & Roy, Deb. 2009. Exploring word learning a high-density longitudinal corpus. *Proceedings of the Thirty-First Annual Conference of the Cognitive Science Society*. (https://dspace.mit.edu/handle/1721.1/66701.)

Sapién, Racquel-María & Hirata-Edds, Tracy. 2019. Using existing documentation for teaching and learning endangered languages. *Language and Education* 33. 560–576. https://doi.org/10.1080/09500782.2019.1622711

Schwartz, Richard & Terrell, Brenda. 1983. The role on input frequency in lexical acquisition. *Journal of Child Language* 10. 57–64. https://doi.org/10.1017/S0305000900005134

Segbers, Jutta & Schroeder, Sascha. 2016. How many words to children know: A corpus-based estimation of children's total vocabulary size. *Language Testing* 34. 297–320. https://doi.org/10.1177/0265532216641152

Shneidman, Laura & Arroyo, Michelle & Levine, Susan & Goldin-Meadow, Susan. 2013. What counts as effective input for word learning? *Journal of Child Language* 40. 672–686. https://doi.org/10.1017/S0305000912000141

Shneidman, Laura & Goldin-Meadow, Susan. 2012. Input and acquisition in a Mayan village: How important is directed speech? *Developmental Science* 15. 659–673. https://doi.org/10.1111/j.1467-7687.2012.01168.x

Sorell, C. Joseph. 2012. Zipf's law and vocabulary. In Chapelle, Carol (ed.), *Encyclopedia of applied linguistics*. Oxford: Wiley-Blackwell. https://doi.org/10.1002/9781405198431.wbeal1302

Van de Weijer, Joost. 2002. How much does an infant hear in a day? In Freitas, M. J. (ed.), *Proceedings of the GALA2001 Conference on Language Acquisition*. s.n.

Wasik, Barbara & Hindman, Annemarie & Snell, Emily. 2016. Book reading and vocabulary development: A systematic review. *Early Childhood Research Quarterly* 37. 39–57. https://doi.org/10.1016/j.ecresq.2016.04.003

Weisleder, Adriana & Fernald, Anne. 2013. Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science* 24. 2143–2152. https://doi.org/10.1177/0956797613488145

Weizman, Zehava & Snow, Catherine. 2001. Lexical output as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. *Developmental Psychology* 37. 265–279. https://doi.org/10.1037/0012-1649.37.2.265

Xu, Dongxin & Yapanel, Umit & Gray, Sharmi. 2009. Reliability of the LENA language environment analysis system in young children's natural home environment. (Technical Report No. LTR-05-2.) Boulder: LENA Foundation.

Yang, Charles. 2016. *The price of linguistic productivity: How children learn to break the rules of language*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9780262035323.001.0001

Zipf, George. 1949. *Human behavior and the principle of least effort: An introduction to human ecology*. New York: Hafner.

William O'Grady
ogrady@hawaii.edu