



Published in final edited form as:

*Commun Disord Q.* 2010 August 1; 31(4): 213–222. doi:10.1177/1525740109358628.

## All-day recordings to investigate vocabulary development: A case study of a trilingual toddler

### Abstract

Major innovations are becoming available for research in language development and disorders. Among these innovations, recent tools allow naturalistic recording in children's homes and automated analysis to facilitate representative sampling. The present study employed all-day recordings during the second year of life in a child exposed to three languages, using a fully-wearable battery-powered recorder, with automated analysis to locate appropriate time periods for coding. This method made representative sampling possible, and afforded the opportunity for a case study indicating that language spoken directly to the child had dramatically more effect on vocabulary learning than audible language not spoken to the child, as indicated by chi-square analyses of the child's verbal output and input in each of the languages. The work provides perspective on the role of learning words by overhearing in childhood, and suggests the value of representative naturalistic sampling as a means of research on vocabulary acquisition.

### Introduction

#### Naturalistic Sampling in Assessment of Language Acquisition and Disorders

For many years researchers have been laying foundations for fundamentally new approaches to the study of child development and childhood disorders, including new ways to investigate conversation and language learning (see e.g., Cassotta, Feldstein, & Jaffe, 1964). One of the long term goals of such research has been to make possible naturalistic, all-day recordings in the home and automated analysis of the acoustic information. Without naturalistic recordings, we are restricted to acquiring data in the artificial environment of the clinic or laboratory, and without automated analysis, naturalistic recordings are too unwieldy to utilize practically. Much progress has been made in recent years on various aspects of these problems, some of it related to automatic detection of features of vocalization (Callan, Kent, Guenther, & Vorperian, 2000; Fell, MacAuslan, Ferrier, & Chenausky, 1999; Prud'hommeaux, van Santen, Paul, & Black, 2008) and facial expression (Messinger, Mahoor, Chow, & Cohn, 2009), and some to technologies for naturalistic recording (Buder & Stoel-Gammon, 2002; Johnson, Christensen, & Bellamy, 1976).

This article provides an example of research that is now possible based on these growing foundations. The work takes advantage of a recently developed system allowing all-day recording through a battery-powered device worn by a child, along with automated analysis to locate utterances of the child and those of other speakers (Xu, Yapanel, Gray, Gilkerson, Richards, & Hansen, 2008; Zimmerman, Gilkerson, Richards, Christakis, Xu, Gray, & Yapanel, 2009). The system makes it possible to representatively sample the vocalizations and vocal environment of a child in ways that were previously infeasible. It is important to emphasize that this is a case study, and thus its results should not be generalized to apply to all children or all circumstances of learning. But at the same time it is important not to underemphasize the indications revealed here of the rapidly growing potential for automated, naturalistic investigation of language acquisition in a variety of settings.

The particular focus of the study was not planned in advance of the recordings – instead, the analysis was conducted opportunistically. The author had conducted a set of all-day recordings

over his trilingual daughter's second year of life using the new recording device and automated utterance recognition algorithms. Using the new technologies it was possible to study the role of directedness of language input on vocabulary learning for each of her input languages.

### The Role of Directed Input in Language Development

This case study was developed in response to themes that have long been central in theories of language acquisition. Two well-publicized views differ substantially: One assumes innateness of a specific language faculty with relatively little role for learning (Chomsky, 1966). The other emphasizes learning and environmental influences but does not necessarily discount innate motoric and cognitive abilities of infants that may predispose humans to be able to learn language and a variety of other complex skills (Tomasello, 2003). These two views differ dramatically in how they portray *the role of language input* in language learning. In the first case language development is assumed to be extremely robust with regard to input types, while in the second, it is emphasized that the form and type of input play important roles in language outcomes.

In the extreme form of the innatist or “nativist” claim, the process of language learning is assumed to be so robust that it is essentially insensitive to subtleties of *how* one speaks to a child. This nativist posture sometimes includes the assertion that language does not even need to be directed to the child in order to produce normal language learning. The assertion is consistent with the extreme Chomskyan idea that language “grows” in the child's mind like any other organ (Anderson & Lightfoot, 2002). As long as the child is exposed to some minimum amount of natural input in the language to be learned, the process will proceed similarly regardless of major differences in how the input is presented. For example Pinker (1994) has claimed that “.in some societies . . . people tacitly assume that children are not worth speaking to, and do not have anything to say that is worth listening to. Such children learn to speak by overhearing streams of adult-to-adult speech...” (p. 155).

In fact very young children can, under some experimental circumstances, learn words by overhearing adult-to-adult speech. The proof of this possibility is based on experiments (Akhtar, Jipson, & Callanan, 2001) where children are taught words in one of two ways: 1) a directed circumstance, where an adult addresses children to teach novel words by demonstration and verbalization, and 2) an undirected circumstance, where the child participates in a non-verbal game with one adult, while another adult uses the directed method of verbal and demonstration instruction to teach novel words to a confederate adult. During the undirected task, the child has the opportunity to listen and watch from a spot nearby in the same room where the silent task is being conducted, and the adult working with the confederate teaches the very same words and uses the very same instructions and demonstrations that are used in teaching children in the studies. Children *do learn words by overhearing*, that is, in the undirected circumstance. However, the overhearing circumstance is a very uncommon one – the child in the experimental overhearing circumstance listens to an adult talking to another adult as if the other adult were a two-year-old, while the observing child is engaged in a silent task. The arrangement may encourage listening to and attending to the overheard interaction *as if the child were a direct participant being spoken to* – after all, the speech produced by the experimenter is structured in a child-directed way, and it is the only speech information in the environment at the time. It is not clear that much of vocabulary learning would ever occur this way, either in Western middle-class circumstances or in any other society.

Another body of research provides proof of passive perceptual learning of word-like sounds by infants. In these studies, infants are presented with acoustic experience in experimental situations, where purely phonological sequences without linguistic meaning are varied systematically (Saffran, Aslin, & Newport, 1996a; Saffran, Newport, & Aslin, 1996b). These perception studies suggest that some components of knowledge required for language

acquisition (phonotactic pattern learning, for example, and recognition of repeating phonological sequences corresponding to words) can be partially acquired based on passive listening and recognition of statistical patterns of input. Meaningful language is not involved, and the meaningless input is presented electronically with no human interactors to indicate how the sounds should be interpreted. A small inventory of syllable types is presented to infants repetitively in an otherwise silent environment. As with the overhearing experiments described above, the circumstance of this passive statistical learning of syllable sequences is not naturalistic, and would seem to correspond to very uncommon experiences in infancy.

To draw the conclusion on the basis of these kinds of research on learning by overhearing, that children do not need to listen to directed speech, or more importantly, *that there is no advantage* to hearing directed speech in the learning of language, would surely be premature and risky. Artificial environments where speech or speech-like stimuli are presented can offer only limited perspectives on what children learn and may need in order to learn in natural environments. Even Pinker acknowledges that children may be helped when their parents speak to them slowly and with purpose (Pinker, 1994). Still, the extreme nativist view has the potential to encourage a parental (or clinical) belief that it simply does not matter whether we speak *to* children or not.

In the functionalist viewpoint there is, in contrast, much emphasis on directed input as a factor in language learning. This viewpoint is heavily influenced by the idea that learning in childhood is facilitated by “scaffolding”, the tendency of parents (and clinicians) to talk to children in ways that are gauged to the children’s level of understanding (Bruner, 1983, 1985). Adults appear sometimes to intentionally simplify language directed to children to a level that is at, or only slightly above, the level of complexity that the child can produce, but this sort of simplification for children has also been claimed to have biological foundations as an intuitive characteristic of parenting (Papoušek & Papoušek, 1987). The idea of scaffolding is often traced back to the notion of a “zone of proximal development” (Vygotsky, 1934) posited to be optimal for teaching children new words or other skills. The zone of optimal input changes of course as the child acquires new skills, and the parent, teacher or clinician is seen as naturally adapting to the child’s progress by presenting more complex input at each new stage.

The notion of scaffolding and the importance of input is supported in a decades-long history of work on “motherese” or “parentese” (Fernald, 1992; Snow, 1972) and on the apparent role it plays in encouraging vocal learning (Huttenlocher, Haight, Bryk, Selzer, & Lyons, 1991; Ninio, 1992). Additional research shows a relation of socio-economic status to amount of talk in families, with apparent consequences for child language learning (Hart & Risley, 1995; Hoff-Ginsberg, 1991). A functionalist overview of empirical outcomes suggests, then, that it may indeed matter how much people talk *to* children. Even the research on passive learning suggests that infants recognize meaningless syllable patterns better if stimuli presented to them in laboratory circumstances include the prosodic patterns of motherese (Thiessen, Hill, & Saffran, 2005). A key point about the nature of most effective input for language learning may be that it needs to be comprehensible (Krashen, 1985), which is to say that learners need to be able to do more than just recognize words – they need also to be able to grasp the global meanings of utterances they hear. By simplifying and drawing attention to the things that are said, adults may well be able to aid children in comprehending what is said, and thus to speed learning along (Goldstein & Schwade, in press).

For the practical purposes of parenting, it is important to provide empirical evidence on the nativist and functionalist views about talking to children. Can parents, teachers, and clinicians reasonably expect that one might give up directed language altogether and still have language learning in very young children proceed on a normal schedule? Could parents indeed talk

almost exclusively to each other and rely on this indirect input (and perhaps the television) to supply the child with necessary language learning material?

## Overview of the Study

**On directedness of input and its effect on vocabulary learning**—The circumstance of simultaneous multilingual learning can provide data for an interesting new test of the possible importance of directed speech to children. Some children learning multiple languages experience sharp differences in how much input in each language is directed toward them. This study considers a child who lived in precisely this circumstance during her first two years, with considerable directed input from her Austrian mother in native German and a smaller but still significant amount of directed input from her Latin American governess in native Spanish. She heard English consistently, but it was primarily in the form of talk between the parents, not directed to her. The father (the author of the paper) speaks German fairly well, and certainly well enough to talk with a one or two-year-old comfortably, but when speaking with his Austrian wife, he spoke overwhelmingly in English, as did she, and these adult to adult conversations were often overheard by the daughter.

**On new recording and automated analysis technology to aid representative, naturalistic sampling**—The study to be reported here tests the question of directedness of input in a single case study of this simultaneous multilingual learner. The importance of the report is heavily methodological, because the procedures that have allowed this effort to be completed are based on technology that affords convenient and effective approaches to representative sampling of both language input to the child and language output from the child, approaches that are only recently accessible and practical to use.

The present case study has, then, two goals:

1. To illustrate the feasibility of all-day recording with automated analysis of vocal activity levels in providing the basis for representative sampling in studies of early language development and potentially in studies on language disorders; and
2. To evaluate the role of directedness (to the child) of linguistic input on expressive vocabulary learning in this child exposed to three languages. It is hypothesized that the representatively selected samples of data from the recordings will show that the child's vocabulary in German and Spanish will be relatively high, corresponding to the fact that the input to the child in those languages was predominantly directed to her. Conversely it is expected that the child's vocabulary in English will be comparatively low, corresponding to the fact that the input in English was primarily *not* directed to her.

## Method

### Participant and Language Environment

The child was the author's daughter, raised as an only child with her mother (native language, German), her author-father (native language, English), and a part-time governess (native language, Spanish). The mother is a very competent speaker of English, and the father a very competent speaker of both Spanish and German. The Spanish-speaking governess was engaged specifically to ensure that the child would learn Spanish, and it was decided that both parents would speak German to her, assuming that the child would acquire English later in the American environment. Conversation between the parents was, however, routinely conducted in English, even with the child present. The child had no known cognitive, hearing or language disability.

## Recording Device and Procedure

The LENA recording device weighs about 70 grams, and is about the size of a package of mints (Christakis, Gilkerson, Richards, Zimmerman, Garrison, Xu, Gray, & Yapanel, 2009; Warren, Gilkerson, Richards, & Oller, in press; Zimmerman *et al.*, 2009). It can be snapped into the chest pocket of a vest or other specially designed clothing for children. The device allows recording for up to 16 hours. Adults can turn it on or off by holding the record button down for several seconds. It was never turned off by the child (it would be difficult for a child to do so). The device was always out of sight and was seemingly unobtrusive while being worn – the child seemed to ignore it. The system records acoustic data at a 16 kHz sampling rate, through a single microphone, 7–10 cm from the infant or child's mouth while in the specially designed clothing. The recording quality is good in circumstances of low noise, but signal-to-noise ratio can be much poorer if there is a high ambient noise level, if there are multiple voices or other kinds of sounds (including television or other electronic sounds) occurring simultaneously in the environment. Recording quality is also affected negatively if anyone (including the child) creates friction noise by touching the area on the clothing where the recorder/microphone is housed.

After each recording was complete, the recorder was connected by USB port to a laptop computer that included the LENA analysis software, which automatically uploaded the recorded data (erasing it from the recorder in the process) and processed it, yielding a count of adult words and child vocalizations. The data could be displayed automatically in the form of histograms of hour by hour adult word and child vocalization counts. Each hour could be displayed (by clicking on the histogram bar for the hour) in another histogram of the word and vocalization counts broken down into 12 five-minute intervals.

On 9 of the 11 recording days, the child woke and was dressed in the special clothing after which the fully-charged recorder was turned on and snapped in place in the chest pocket, where it could not be seen. She would wear the clothing with the device in it all day. At bedtime or nighttime bath time the device was removed and turned off. On two of the recording days, the recording period was much less than the full day because the author of the paper, who has the habit of toying with options in software, had mistakenly made software adjustments that had on two occasions disabled the procedure that would normally have erased the recorder's data bank with each upload. Not having noticed the failure to erase, two recordings were begun with a recorder that was already almost full. As a result, these two days produced very short recordings (0.53 and 3.28 hours).

The activities occurring in the recordings were the normal ones of the household, mostly involving the child playing with one or another of the caregivers, eating with one or more of the caregivers, and so on. Sometimes she would go with one of the caregivers to a nearby park to play, to take a walk in the quiet neighborhood where she lived, or to the nearby grocery store. In short, the recordings were made in the natural environment in which the child lived.

## Selection of Days for Recording

The study began simply as an asystematic attempt to document daily vocal activities of the household, with the intent to sample all the sorts of language input that the child received. Recording days were varied to ensure representation of all three languages, with considerable representation of input from each of the three primary caregivers. At the same time, no specific research questions were at stake at the time the recordings were being made, and they were scattered across the approximately one-year period from when the child was 11 to 24 months of age. The average length of the 11 recordings was 9.8 hours (11.56 without the two short ones). One recording occurred at 11 months, one at 13, and 9 (including the two short ones) occurred between 19 and 24 months.

## Representative Sampling of Input in Each Language

The decision to conduct a study on the role of directed input developed during the period of recording as it occurred to the author that the child's learning situation was somewhat peculiar. She was clearly hearing a good deal of English, but rarely was it spoken to her. The recordings and automated analyses provided a rich indication of how much talk was going on during each recording, but the automation did not offer any information about what language was being spoken by the child or the adults. It was possible, however, to listen selectively to five-minute samples of any recording based on the LENA record in order to determine what speakers were present and to count words and/or transcribe utterances of both child and adults, and thus to determine which language was being spoken.

One goal of this study was to illustrate the convenience of representative sampling with this method. 39 five-minute periods were selected from across all the recordings, including multiple time periods with all the primary caregivers and combinations of them. The selections from each day were chosen based on the automatic histogram reports on vocalization rates from hours during which child vocalization rate was relatively high, and during which the rate was also high in the five-minute periods.

## The Word Count Procedure

Words in these samples were counted by the author alone, because a competent speaker of all three languages was required for the task. Each five-minute sample was located through the LENA software, from where a waveform window could be opened with a single mouse click, and the waveform sample could be played and paused as counts were made. A pencil and paper recording sheet was used for each sample, with rows corresponding to speakers and columns to word tokens produced in each language-related category. The listener checked the sheet in the appropriate cell each time he heard that a word (a token of any word) had been produced by the child or any other speaker.

Five language-related categories were used for each word produced: Spanish, German, English, Ambiguous or Unintelligible. Ambiguous words were not language-specific. They included, for example, names of people, pronounced similarly and often indistinguishably in any of the languages, and other child words that were used equivalently across languages – for example, the word “baba” was used by all the caregivers and the child in all three languages to mean “bottle”. Unintelligible words were also common both from the child and the adults because of the noise that was commonly present in the recordings due to household events or movement of the child, and due to the child's limited phonological skills.

Further, a record was kept of *word types* produced by the child: Each time a new word (one that had not been previously noted in the coding) was spoken by the child, it was listed by language so that a record not only of word tokens spoken by the child in each language, but also of word types could be reported. Finally each of the five columns was subdivided for the adult words into words spoken to the child (directed to child) and words spoken to anyone else (not directed to child).

Intraobserver reliability for number of words spoken for the various speakers in each language was estimated empirically by a second coding of two of the samples where all three primary caregivers were present, weeks later, long after the numbers associated with the samples had been forgotten. The listener counted 424 (mean = 15.1, sd = 22.2 over the 28 categories of person and language) words on the first coding and 435 (mean = 15.5, sd = 23.6) on the second. The correlation for the 28 comparisons of numbers of words counted in each category (for example, Mother's English words, Father's Spanish words, etc.) in the first coding with the numbers of words counted in each category in the second coding was 0.96. The average number

of words counted in each category for the reliability test was 15.1 words, and the average absolute value of the difference between the 28 values representing each category counted was 3.5 words. The difference is not surprising because of the noisiness of the tapes. Still the correlation suggests stability of the observer's ability to recognize basic patterns of production in each of the languages.

### Tabulation of the Word Counts

The data on how many words the child used in each language and on how many she heard in each language, both directed and undirected, are analyzed below based on two different tabulations. The first way is simple, being based on raw counts only – no changes were made, and the raw numbers of child and adult words were simply entered in a 3×3 table (x axis = Spanish, German, English; y axis = words directed to child, words not directed to child, and words produced by the child, see Tables 2 and 3) based on the sum of counts from all the 39 five-minute samples.

This raw-count tabulation ignores the possibility that the 39 selected periods may not have been representative of the real-life occurrence of the primary caregiving circumstances determining the child's input across the year of the study. So an additional method (the rebalanced tabulation) was also used to adjust the data in accord with the distribution of the caregiving circumstances. First estimates were derived based on recollections of the parents and written records that had been kept by the parents providing empirical information on how often each of the circumstances had occurred during the year of sampling (see Table 1).

Then each of the 39 samples was assigned to one of the caregiving circumstances based on which adults were actually present in each five-minute period, and the counts in each cell of the 3×3 table for that period were multiplied by the proportions in Table 1 to yield the rebalanced tabulation, taking into account the estimated distribution of caregiving circumstances. Finally each cell of the rebalanced tabulation was multiplied by a correction factor to produce a proportionally unaltered rebalanced tabulation where the sum of all cells was equal to the sum of cells in the raw-count tabulation.

The results to be reported below yield all the same conclusions and significant results whether the raw-count tabulation or the rebalanced tabulation was used.

## Results

With regard to the first goal of the study, the all-day recordings and automated analysis provided a very workable method of sampling from the naturalistic language environment of the child. Recordings were made without difficulty; processing to obtain the day by day, hour by hour, five-minute by five-minute reports was uneventful, requiring only to connect the recording device to the computer housing the processing software; and location of time periods of relatively high vocal activity was easily managed by simply observing and clicking on the automatically produced bar charts of word counts and vocalizations. From that point, coding could be conducted on the samples in the same way one might work from a digital tape recorder.

With regard to the second goal, the data showed extremely strong patterns confirming the predictions that the child's output words would be high in the languages primarily spoken to her, and low in the language that was primarily *not* spoken to her, but instead spoken between adults in her presence. The raw numbers of words produced in each language for both the adults and the child as indicated by the observations are presented in Figure 1. A total of 9119 words were counted. Excluding the ambiguous and unintelligible items, 67 per cent of the counted input words were German. Note also that input directed to the child tended to be enormously more common than input not directed to the child in German and in Spanish, while in English

most of the input sampled was not directed to the child. At the same time, the figure shows that the child produced very little English. Less than 4 per cent of all word tokens spoken by the child were in English even though 17 per cent of the input (combining directed and undirected) was in English.

Table 2 provides the 3×3 table of data based on the raw-count tabulation of words (tokens) as well as the rebalanced tabulation of words that took into account the estimated amounts of time the child had spent in each of the caregiving circumstances (as indicated in Table 1).

Table 2 provides further unambiguous indications that the predictions of goal two were confirmed, namely that the child used fewer words in English than would have been predicted by the amount of input (both directed and undirected) that she experienced in English, and conversely, that she used many more words in both German and Spanish than would have been predicted based on the input values. The two 3×3 matrices (one for raw counts and one for rebalanced counts) in Tables 2 were subjected to chi-square analysis to illustrate the massively significant effects of the study. For both the raw counts and the rebalanced counts, the chi-square value exceeded 3000, corresponding to an extremely low probability of non-independence of the input and output counts for the three languages ( $p < 10^{-30}$ ). The statistical indication can be interpreted to mean that the child's tendency to use words in each of the three languages and the adults' tendencies to direct words in the three languages were extremely *interdependent*. The sources of this interdependence can be seen by referring to the adjusted residual values (Bakeman & Gottman, 1997) for each cell in the two matrices as recorded in Table 3.

Adjusted residuals are z-scores representing the size of the discrepancy between the observed distribution of words in the nine cells of the table and the expected distribution derived through the chi-square formula, the discrepancy being expressed in standard deviation units. To understand the adjusted residuals consider an example: The child produced 45 words in English (raw count). Had this value been independent of the input pattern in the three languages, one would have expected (based on the chi-square formula) the child to produce 186 words in English. Based on the bottom right hand cell of the Table 3, it can be seen that the observed value of 45 words in English was more than 12 standard deviations *below* the expected level of 186. Similarly the adjusted residuals indicate that the child's input in English was massively imbalanced in terms of directedness – undirected words of English input occurred 56 to 60 standard deviations *above the level expected*, and directed words occurred 30 to 34 standard deviations *below the level expected*. Moreover, the data confirm that words not directed to the child occurred at far lower rates than would have been predicted for both Spanish and German. The child's input pattern was starkly clear. The English she heard was predominantly spoken between adults, while the Spanish and German she heard was predominantly spoken to her.

Figure 2 provides a summary on word *type* and *token* counts for the child, comparing the proportions of occurrence in the three languages. Here the rebalanced token counts for the child are presented, along with the sum of the rebalanced directed and undirected adult input words in each language. The data show that for types as well as tokens, the child produced very little English. In addition the data suggest that the child's Spanish included more types than would have been predicted by the number of tokens produced in Spanish, with the reverse being true for German.

To summarize, the data show that the directedness of input words in the three languages was strongly predictive of both the number of tokens and the number of types of words that the trilingual child used in each of the three languages. Consider these indicators of powerful effect sizes.



1. The child's samples showed nearly 14 times as many lexical items (types) in Spanish as in English, even though the total amount of input (number of words spoken by adults based on either raw counts or rebalanced counts) was comparable in the two languages. What was different, was that words spoken as input in Spanish were overwhelmingly directed to the child, while the opposite was true of English. Similarly, the proportion of word *tokens* spoken in Spanish by the child was nearly five times higher than in English based on the rebalanced tabulations.
2. The child showed nearly 25 times as many lexical items in German as in English. In German the proportion of word tokens spoken by the child after rebalancing was relatively similar (0.78) to the proportion of *directed* input in German (0.83). The same relative proportional similarity was seen in Spanish (child tokens = 0.18; directed input = 0.12). In both German and Spanish, however, the proportion of word tokens spoken by the child exceeded the *undirected* input proportion by a factor of more than 3.5.
3. In contrast, the proportion of *undirected* input in English was 22.5 times higher than the child's proportion of English output in word tokens.

## Discussion

In the methodological domain, a primary result in the present study concerns the application of naturalistic all-day home recording with automated analysis to locate utterances of children and adults. Collecting the data in this study was relatively simple. The home recordings required only charging the device the night before, having the appropriate clothing available and clean in the morning, and then turning on the device and placing it in the chest pocket of the clothing at the time of dressing the child. Automated analysis only required plugging the device into the computer.

The present study presents the first trial of a scheme for representative sampling based on such naturalistic recordings. Adequate sampling for a multilingual study such as this one requires forethought to ensure that all the relevant caregiving circumstances (which determine the input languages) have been recorded. Then a scheme of rebalancing for caregiving circumstance can be developed based on how often each of those circumstances occurred – of course if the recordings are sufficiently numerous, they could themselves provide a basis for determining the distribution of caregiving circumstances. Then sampling from the all-day recordings can focus on periods of time from all the types of circumstances and a multiplier can adjust the data to make them represent the proportion of time the child experienced each one.

Obviously the scheme for representative sampling could be made more elaborate and sophisticated than in this first and only partially planned attempt. But even at this preliminary level, it is reasonable to assert that the study provides a more representative picture of the child's production vocabulary in each of three languages than could have been acquired by any practicable laboratory method. Laboratory time is simply too costly, and even with many hours of recording in the laboratory, we have poor assurance that the outcome would be representative of the child's general communicative performance.

Without these conveniences, the study would not have been practical for the author to conduct. Not only was it important that it be easy to acquire the recordings and the automated analyses, but also it was important that coding time could be kept to a minimum. Only by representatively sampling relatively small numbers from a large recording corpus was it possible to keep the coding time low. While this is just a beginning, it illustrates that naturalistic representative sampling can now be approached with tools that are convenient and practical. Substantial scientific as well as clinical applications can be anticipated (Oller, Gray, Richards, Gilkerson,

Warren, & Niyogi, 2009; Warren *et al.*, in press; Xu, Gilkerson, Richards, Yapanel, & Gray, 2009).

In the empirical domain, the study provides reason to doubt the strong nativist view espoused by Pinker and others, the view that language learning is so robust and so deeply innate, that learning will proceed normally whether we talk *to* our children or not. Of course this is a case study, but its results are unambiguous in indicating that in this single multilingual circumstance, it mattered greatly whether speech was addressed to the child or to another adult. The results do not overturn the indication that very young children may learn *some* words through overhearing. Especially as children get older, it seems likely that instances of learning by overhearing may occur commonly. Moreover, the results of the present study should not be assumed to apply to every cultural circumstance equivalently. The outcome should be interpreted to mean instead that overhearing of adults talking to each other about adult matters played at most a very small role in vocabulary learning in this child in the second year of life compared with the very strong role played by speech directed to the child.

Lieven (1994) provides a review of cross-cultural empirical data on language addressed to children. The results do suggest strong differences across cultures in how adults speak to children, but they do not provide a basis for the conclusion that adult-to-adult speech is the *primary* source of information used by children in some cultures to learn to talk. Lieven's review suggests at least three reasons to withhold such a conclusion: 1) Little is known for cultures where adults interact verbally relatively little with infants and very young children about how much speech is directed to infants and young children by other (older) children; 2) Adults talking to somewhat older children (rather than to other adults) may also supply a substantial (overheard) source of input to infants and very young children; and 3) Research claiming that there exist cultures where adults do not talk to infants and very young children is largely qualitative, and it seems likely that careful quantitative observations will produce a more nuanced view of the matter, with only quantitative differences across cultures in how often speech is addressed to very young children.

The present research offers encouragement for the supposition that directedness plays a very important role in vocabulary learning, and it offers a cautionary note regarding the claim that children may be able to learn vocabulary exclusively or primarily from overhearing adult to adult speech. It suggests specifically that this child, who was given the choice of listening to speech directed to her or to speech not directed to her, focused her attention strongly on the former, and thus learned much more from directed than undirected input. This pattern could be particularly associated with multilingual learning. For cultures where little speech is directed to them, children may be required to find other bases for selective attention in learning.

## Applications for Practice

Of course the question of directedness of language input has important clinical implications for cases of delayed or disordered language development, both regarding what one might choose to do in schools or in therapy and regarding what one might wish to advise parents to do at home. The results of the present study do not address the question of whether overheard speech (for example speech between adults) might be helpful for the one-year-old child's learning process, especially if it is structured to be comprehensible by the child. The results offer the simple suggestion that directed speech appears to have far more impact on learning than undirected speech in the naturalistic circumstance that was sampled here.

While one-year-olds or even younger infants may notice some aspects of speech even if they do not actively attend to it (Saffran *et al.*, 1996a; Saffran *et al.*, 1996b), there is little reason to believe they comprehend the general thrust of most adult-to-adult conversation. The child in the present study seemed to tune out during adult-to-adult conversations, which tended to be

about such matters as how to schedule the next dinner party, who was going to do the shopping for the evening, and the condition of the roof. The child may have learned a little about English by overhearing these conversations, but it appears that what she learned about Spanish and German from directed speech was enormously more potent, presumably because it was directed to her, and it was gauged to her level of understanding. The fact that she seemed to have attended to and learned to use much more Spanish and German than English may well have been simply the product of her having been given the opportunity to attend to comprehensible input, and presumably input that was relevant to her, in Spanish and German (Krashen, 1985).

This research suggests that clinicians working with children who are at risk or who have a language disorder should preserve and expand methods designed to attract children's attention through directed communications during therapy. It suggests that such directed communication provides precisely the kind of material that has the potential to produce vocabulary learning. Furthermore, the research suggests that parents who talk to their children regularly, and who generally try to engage their children directly, are doing precisely what they ought to be doing.

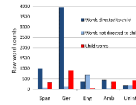
## Acknowledgments

This research was supported in part by the Plough Foundation and in part by a grant from the National Institute of Child Health and Human Development, R01 HD046947 to D. Kimbrough Oller, PI. The author is an unpaid member of the Scientific Advisory Board of the not for profit LENA Foundation, which was established in February, 2009, at which time it acquired all the assets of Infoture, the prior for profit company that originally developed the recorder and processing software used in this study. The author was an original member of the Scientific Advisory Board of Infoture and was paid occasional consultation fees for that role from 2004 through early 2008.

## References

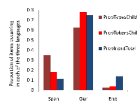
- Akhtar N, Jipson J, Callanan MA. Learning words through overhearing. *Child Development* 2001;72(2): 416–430. [PubMed: 11333075]
- Anderson, SR.; Lightfoot, DW. *The language organ: Linguistics as cognitive physiology*. Cambridge, UK: Cambridge University Press; 2002.
- Bakeman, R.; Gottman, JM. *Observing interaction: An introduction to sequential analysis*. 2. Cambridge: Cambridge University Press; 1997.
- Bruner, J. *Child's talk*. New York: Norton; 1983.
- Bruner, J. Vygotsky: A historical and conceptual perspective. In: Wertsch, JV., editor. *Culture, communication and cognition: Vygotskian perspectives*. 1985. p. 21-34.
- Buder EH, Stoel-Gammon C. Young children's acquisition of vowel duration as influenced by language: Tense/lax and final stop consonant voicing effects. *Journal of the Acoustical Society of America* 2002;111:1854–1864. [PubMed: 12002869]
- Callan DE, Kent RD, Guenther FH, Vorperian HK. An auditory-feedback-based neural network model of speech production that is robust to developmental changes in the size and shape of the articulatory system. *Journal of Speech, Language, and Hearing Research* 2000;43:721–736.
- Cassotta L, Feldstein S, Jaffe J. AVTA: A device of automatic vocal transaction analysis. *Journal of the Experimental Analysis of Behavior* 1964;7(1):99–194. [PubMed: 14120152]
- Chomsky, N. *Cartesian linguistics*. New York: Harper and Row; 1966.
- Christakis D, Gilkerson J, Richards J, Zimmerman F, Garrison M, Xu D, et al. Audible TV is associated with decreased adult words, infant vocalizations, and conversational turns: A population based study. *Archives of Pediatrics and Adolescent Medicine* 2009;163:554–558. [PubMed: 19487612]
- Fell HJ, MacAuslan J, Ferrier LJ, Chenausky K. Automatic babble recognition for early detection of speech related disorders. *Journal of Behaviour and Information Technology* 1999;18(1):56–63.
- Fernald, A. Human maternal vocalizations to infants as biologically relevant signals: An evolutionary perspective. In: Barkow, JH.; Cosmides, L.; Tooby, J., editors. *The adapted mind: Evolutionary psychology and the generation of culture*. Oxford: Oxford University Press; 1992. p. 345-382.

- Goldstein, MH.; Schwade, JA. From birds to words: Perception of structure in social interactions guides vocal development and language learning. In: Blumberg, MS.; Freeman, JH.; Robinson, SR., editors. *The oxford handbook of developmental and comparative neuroscience*. Oxford, UK: Oxford University Press; in press
- Hart, B.; Risley, TR. *Meaningful differences in the everyday experience of young American children*. Baltimore: Paul H. Brookes; 1995.
- Hoff-Ginsberg E. Mother-child conversation in different social classes and communicative settings. *Child Development* 1991;62:782–796. [PubMed: 1935343]
- Huttenlocher J, Haight W, Bryk A, Selzer M, Lyons T. Early vocabulary growth: Relation to language input and gender. *Development Psychology* 1991;27:236–248.
- Johnson SM, Christensen A, Bellamy GT. Evaluation of family intervention through unobtrusive audio recordings: Experiences in bugging children. *Journal of Applied Behavioral Analysis* 1976;9(2):213–219.
- Krashen, SD. *The input hypothesis: Issues and implications*. London: Longman; 1985.
- Lieven, EVM. Crosslinguistic and crosscultural aspects of language addressed to children. In: Gallaway, C.; Richards, BJ., editors. *Input and interaction in language acquisition*. Cambridge, UK: Cambridge University Press; 1994. p. 56-73.
- Messinger D, Mahoor MH, Chow SM, Cohn JF. Automated measurement of facial expression in infant–mother interaction: A pilot study. *Infancy* 2009;14(3):285–305. [PubMed: 19885384]
- Ninio A. The relation of children’s single word utterances to single word utterances in the input. *Journal of Child Language* 1992;19:291–313.
- Oller, DK.; Gray, S.; Richards, JA.; Gilkerson, J.; Warren, SF.; Niyogi, P. Evaluating naturalistic home recordings: Automated analysis of acoustic features of infant sounds. Denver, CO: Biennial Meeting of the Society for Research in Child Development; 2009.
- Pinker, S. *The language instinct*. New York: Harper Perennial; 1994.
- Prud’hommeaux ET, van Santen J, Paul R, Black L. Automated measurement of expressive prosody in neurodevelopmental disorders: International Meeting for Autism Research. 2008
- Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. *Science* 1996a;274:1926–1928. [PubMed: 8943209]
- Saffran JR, Newport EL, Aslin RN. Word segmentation: The role of distributional cues. *Journal of Memory and Language* 1996b;35:606–621.
- Snow C. Speech to children learning language. *Child development* 1972;43:549–565.
- Thiessen ED, Hill EA, Saffran JR. Infant-directed speech facilitates word segmentation. *Infancy* 2005;7(1):53–71.
- Tomasello, M. *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press; 2003.
- Vygotsky, LS. *Thought and language*. Hanfmann, E., translator. Cambridge, MA: MIT Press; 1934.
- Warren SF, Gilkerson J, Richards JA, Oller DK. What automated vocal analysis reveals about the language learning environment of young children with autism. *Journal of Autism and Developmental Disorders*. in press.
- Xu, D.; Gilkerson, J.; Richards, JA.; Yapanel, U.; Gray, S. Child vocalization composition as discriminant information for automatic autism detection. Minneapolis: IEEE Engineering in Medicine and Biology Society; 2009.
- Xu, D.; Yapanel, U.; Gray, S.; Gilkerson, J.; Richards, J.; Hansen, J. Signal processing for young child speech language development. Paper presented at the 1st Workshop on Child, Computer and Interaction; Chania, Crete, Greece. 2008.
- Zimmerman F, Gilkerson J, Richards J, Christakis D, Xu D, Gray S, et al. Teaching by listening: The importance of adult-child conversations to language development. *Pediatrics* 2009;124:342–349. [PubMed: 19564318]



**Figure 1.**

Raw word counts for each of three languages (Spanish, German and English) spoken to the child (dark blue), in the presence of the child but not directed to her (light blue), and by the child (red). Words that could not be assigned to a language were termed “ambiguous” (Amb). Words that could not be identified were termed “unintelligible” (Unint). The figure reports on 9119 words counted from 39 five-minute samples (5934 directed to the child, 1108 not directed to the child, and 2077 produced by the child herself). 7483 of the words were produced intelligibly in one of the three languages.



**Figure 2.** Proportions of words that occurred in each of the three languages in terms of word types produced by the child (rust color), word tokens produced by the child (red color), and identifiable words spoken by adults in one of the three languages (termed “input”, blue color). Word tokens produced by the child and words spoken by the adults are based on the rebalanced tabulations reflecting estimated actual time spent by the child in each of the caregiving circumstances (see Table 1).

**Table 1**

Estimated proportions of time the child spent in each of the caregiving circumstances across the period of sampling

Mother and father with the child	0.285
Mother alone with the child	0.276
Father alone with the child	0.156
Governess alone with the child	0.155
Other circumstances	0.128

**Table 2**

Total raw word counts/rebalanced word counts, excluding ambiguous and unintelligible items

	Spanish	German	English	Total
Words directed to the child	1003/668.5 <sup>a</sup>	3955/4466.2	358/223.0	5316/5357.7
Words not directed to the child	70/42.9	136/171.3	686/645.4	892/859.7
Words spoken by the child	328/231.1	902/985.4	45/49.0	1275/1265.7
Total	1401/942.6	4993/5623.0	1089/917.4	7483/7483 <sup>b</sup>

<sup>a</sup>The values for the rebalanced counts are not whole numbers because they are estimates in accord with the rebalancing procedure, with cell values adjusted from the raw counts based on the estimated distribution of caregiving circumstances (see Table 1).

<sup>b</sup>The grand total is, by design, identical for the raw counts and rebalanced counts – rebalancing redistributed the total raw count across cells.



**Table 3**

Adjusted residuals from chi-square analysis: Based on raw counts/rebalanced counts

	Spanish	German	English
Words directed to the child	0.50/-0.49	22.06/26.11	-30.04/-33.91
Words not directed to the child	-8.87/-7.14	-34.77/-39.82	56.27/59.69
Words spoken by the child	7.04/6.67	3.34/2.46	-12.26/-9.98