The LENA^{TM} system applied to Swedish: Reliability of the Adult Word Count estimate

Iris-Corinna Schwarz[1], Noor Botros[2], Alekzandra Lord[2], Amelie Marcusson[2], Henrik Tidelius[2] and Ellen Marklund[1]

[1]Stockholm Babylab, Phonetics Laboratory, Dept. of Linguistics, Stockholm University, Sweden
[2]CLINTEC, Speech Pathology Unit, Karolinska Institutet, Stockholm, Sweden

iris-corinna.schwarz@ling.su.se, ellen.marklund@ling.su.se

Abstract

The Language Environment Analysis system LENA™ is used to capture day-long recordings of children's natural audio environment. The system performs automated segmentation of the recordings and provides estimates for various measures. One of those measures is Adult Word Count (AWC), an approximation of the number of words spoken by adults in close proximity to the child. The LENA system was developed for and trained on American English, but it has also been evaluated on its performance when applied to Spanish, Mandarin and French. The present study is the first evaluation of the LENA system applied to Swedish, and focuses on the AWC estimate. Twelve five-minute segments were selected at random from each of four day-long recordings of 30-month-old children. Each of these 48 segments was transcribed by two transcribers, and both number of words and number of vowels were calculated (inter-transcriber reliability for words: r = .95, vowels: r = .93). Both counts correlated with the LENA system's AWC estimate for the same segments (words: r = .67, vowels: r = .66). The reliability of the AWC as estimated by the LENA system when applied to Swedish is therefore comparable to its reliability for Spanish, Mandarin and French.

Index Terms: parental speech input, parent-child interaction, LENA system, Swedish

1. Introduction

The amount of speech exposure is a simple measure, but yet central for children's language acquisition. Children with large vocabularies and rapid vocabulary growth are more likely to have mothers who use a high number of words compared to children with smaller vocabularies and slower vocabulary development rate [1]. This relationship between more parent speech input and larger child vocabularies has been shown many times over since this first classical study [2, 3, 4]. Many of these studies have focused on English-learning children and their families, but not all. For example, the number of utterances that Spanish-speaking mothers address to their Spanish-learning child at 18 months correlates with child vocabulary size at 24 months [3]. Importantly, the effect of amount of speech input on child vocabulary development is found only for child-directed speech, not for conversations between adults simply overheard by the child [2].

Of course, the amount of child-directed speech, quantified by number of words, number of utterances, or duration, is a very basic measure. There are other factors in the speech input, such as lexical richness and syntactic complexity [4] and for example verb diversity [5], that are directly related to language development. However, despite their relative simplicity, measures of speech input amount are consistently found to be reliable predictors of children's later language outcomes [e.g., 2, 3, 4]. On the basis of these consistent findings, researchers have issued recommendations for parents to speak more with their children [6, 7, 8].

There is clearly value in using the simple measure of speech input amount, both in research and clinical practice. However, a methodological bottleneck in assessing speech input amount is manual transcription and/or markup of the recordings. This is where the Language Environment Analysis system LENA (LENA Research Foundation) can be useful. It has been developed with the express purpose of estimating the amount of speech present in the auditory environment of children.

The LENA system consists of a patented hardware recording unit, the Digital Language Processor (DLP; version 2.18.00, for general technical specifications see [9]) and an analysis software program called LENA-Pro (V3.4.0-143r11780, LENA Research Foundation, Boulder, CO, USA). The software is based on an acoustic model for automatic speech recognition that as a first step identifies human speech among other audio signals [10]. In a second step, it further subdivides the signal into eight categories. The segments of human speech are separated into 1) target child speech (identified by proximity to recording device and child age), 2) other child speech (identified by distance to recording device and child age), 3) female adult speech, 4) male adult speech, and 5) overlapping speech. The non-speech segments are separated into 6) electronic media (e.g. TV, radio, tablets), 7) noise (essentially all non-identifiable sounds) and 8) silence.

Based on these segmentations, the software can then estimate 1) Adult Word Count (AWC; the approximated number of adult words spoken in close proximity to the target child), 2) Conversational Turns (CTC; number of instances that either the child or the adult speaks and is responded to by the other within five seconds), and 3) Child Vocalization Count (CVC; the number of non-vegetative vocalizations of the target child surrounded by at least 300 ms of vocal pause).

LENA's AWC estimate has been evaluated against human listeners' word counts for American English. The recordings were listened to in short segments, and for each segment the listeners noted how many words they heard. This original evaluation was based on a tap-counted pre-selection of segments that contained large amounts of near and clear speech according to the LENA system. The system's AWC estimate correlated significantly with the human listeners' count (r = .92, p < .01) [11]. Similar evaluations, but including transcribing the recorded speech and counting the transcribed words, show a somewhat smaller correlation between human word counts and the LENA system's AWC, ranging between r= .71 and r= .85 [12, 13, 14].

The LENA system has also been evaluated when applied to languages other than American English. The reliability of the AWC estimate specifically has so far been evaluated for Spanish, Mandarin and French [2, 15, 16]. In the evaluation of the AWC estimate of the LENA system applied to Spanish, 60 minutes out of ten recordings were transcribed, and the words in the transcription counted. The reliability (r = .80) was found to be within the range of that for American English [2]. Similarly, in the Mandarin evaluation, a correlation of r=.73 between transcribed Shanghai dialect words and the estimated AWC was found [15]. When applying the LENA system to French however, the correlation between the system's AWC and human word count was somewhat lower (r = .64) [16].

The present study is the first to report an evaluation of the LENA system applied to Swedish. It focuses on the AWC estimate, aiming to assess its reliability. It is expected that the reliability of the LENA system's AWC will be within the range of those reported in previous evaluations of the LENA system applied to languages other than American English.

2. Method

2.1. Participants

The participants of the present study were all part of an ongoing longitudinal study on parent-child interaction at Stockholm Babylab, Stockholm University. Parents were originally contacted via mail (addresses to newborn children living in the greater Stockholm area were obtained from the Swedish Tax Agency) and agreed to participate in the three-year long study with four lab visits per year, from when their child was three months old. In the third year of the study, they were invited to contribute with day-long recordings in the home environment in addition to the visits to the lab. A subset of the parents agreed, resulting in 24 recordings of the audio environment of 30-month-old children. Out of successfully completed recordings in which both primary caregivers spoke Swedish, four recordings (two boys, two girls) were selected for the present study.

2.2. Recording procedure

Caregivers were instructed on how to use the DLP when they received the device at the lab. They were asked to record at home on a typical weekend day on which they would spend time with their child, but to avoid days with sports events or birthday parties in order to preserve intelligibility of the recorded speech. They were also asked to avoid recording on days when the child or a parent was sick, since those are not likely to reflect the typical environment in terms of adult-child interactions.

After turning the device on in the morning, caregivers were supposed to leave it on until it turns itself off at night. They were instructed on how to insert the DLP into the pocket of a vest, which the child was to wear all day, except for nap and bath time. The vest was worn inside thicker outdoor wear when outside (the recordings took place in early Swedish spring) which reduced audio quality. The families had the opportunity to list time points which they did not want to include in manual analysis, if for example they had discussed sensitive information. Caregivers also agreed to inform any person coming close to the child wearing the DLP that they would be recorded. Families had the possibility to contact on-call research staff for support during their recording day. After the day of the recording, the devices were returned to the lab via courier.

2.3. Data selection procedure

Each of the four recordings in the present study was at least twelve hours long. From each recording, one five-minute segment was selected at random from within each of the first twelve complete hours of recording. The audio files and the LENA system's estimated AWC were extracted from Lena-Pro for each of those 48 segments.

2.4. Transcription procedure

The transcribers used the automatic speech recognition software Dictation for Swedish (built-in feature of Mac OS from Yosemite onwards; Apple Inc., Cupertino, CA, USA). They listened to the discernible parent speech in the audio files, and repeated what was said to the Dictation software, which converted the spoken words to text in a document. Transcribers adjusted speech recognition mistakes in the text as it was being written. The speech was thus orthographically transcribed, and formal spelling of Swedish words was followed, except in cases when the standard spelling and the spoken version of the word differed in number of syllables. For example, the formal spelling of the word *nån* ("someone") is *någon*, but it is rare that both syllables are pronounced. Likewise, the formal spelling of the word "no" would be *nej*, but it is often pronounced as *näe* with two syllables instead of one. In those and similar cases, the informal spelling was used. Nonspeech sounds such as laughter, as well as vegetative sounds such as snorting, aspirated breathing and coughing were not included in the transcriptions. There were four transcribers in total, and each five-minute audio file was transcribed by two different transcribers.

2.5. Measures

Two measures were taken from each dictation transcription: 1) the number of orthographic words, and 2) the number of vowels. The latter was included based on the hypothesis that the LENA system's AWC estimate is possibly – at least in part – based on prosodic cues to syllables in the audio recording. If this is the case, then number of syllables may be a more stable measure to use when evaluating LENA used on languages other than American English, as the average number of syllables per word differs between languages.

The human word and vowel count were both tested for correlation with the LENA system's ACW estimate. Statistical analyses were performed in SPSS 21 (International Business Machines Corp., Armonk, New York, USA).

3. Results

Pearson's correlation coefficients showed very high inter-transcriber correlation both for word counts (r = .95, p < .01) and for vowel counts (r = .93, p < .01; see Figure 1).

A moderate correlation was found between the LENA system's AWC estimate and the transcribers' word counts (r = .67, p < .05) as well as their vowel counts (r = 66, p < .01; see Figure 2).

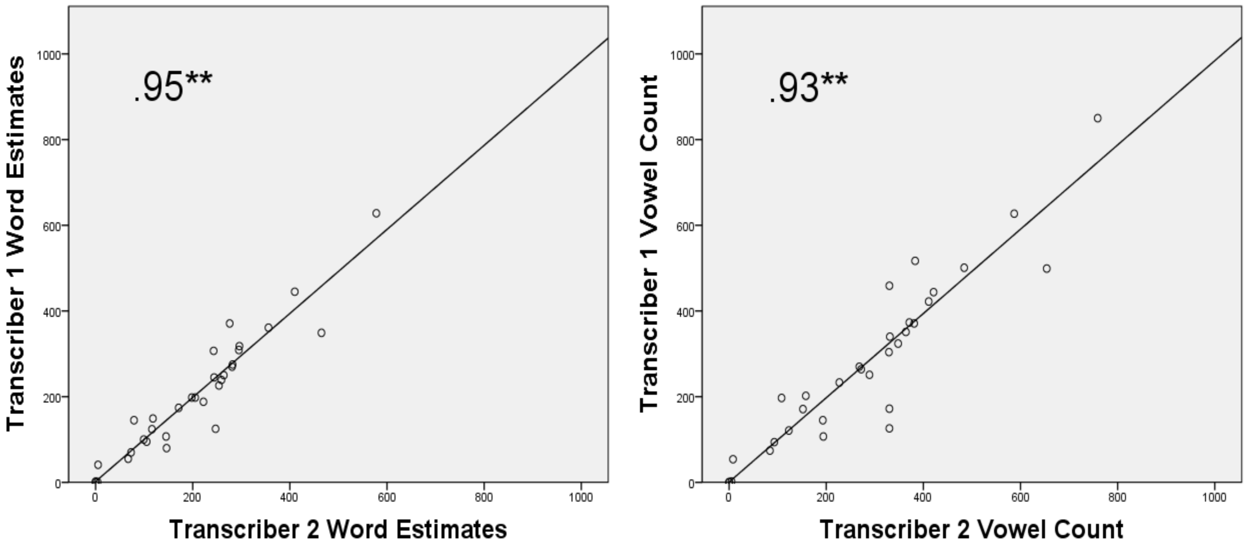

Figure 1: *The inter-transcriber reliability between primary (Transcriber 1) and secondary transcribers (Transcriber 2) as calculated by Pearson's correlation coefficients showed very high correlation both for word estimates and vowel counts.*

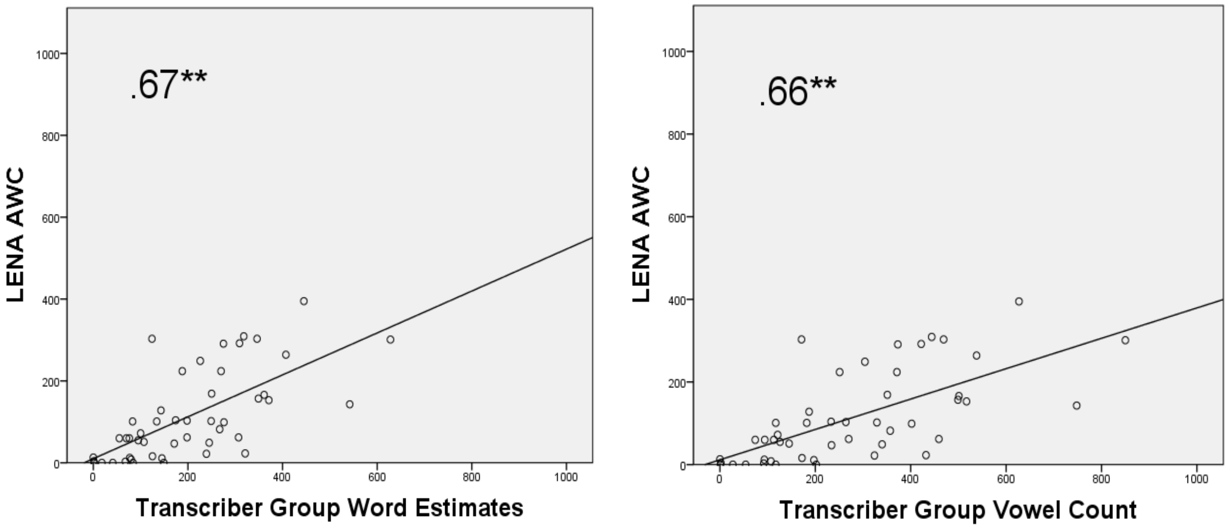

Figure 2: *Pearson correlation coefficients between LENA AWC and the transcriber group show moderate correlations both for word estimates and vowel counts.*

4. Discussion

As expected, the reliability of the AWC estimate in the LENA system when applied to Swedish was at the same general level as those found in other similar studies, albeit in the lower range and most similar to the French reliability result [16].

The present study differs from several of previous evaluations of the AWC measure in terms of how segments used for analysis were selected. In most previous evaluations, segments have been selected based on high amounts of adult and/or child speech (estimated either by the LENA system or by an unspecified estimation procedure) [11, 15, 16]. In the evaluation of LENA applied to Spanish [2], segment selection was based on the AWC estimate, but the selection procedure ensured segments with low AWC were included. In the present study, selection of segments was independent of the amount of speech in the recording: one random five-minute segment was selected per hour of recording, for the first twelve hours of recording. This ensures that the evaluation of the system is not limited to situations in which speech can be detected.

Not surprisingly, the highest reliability of the AWC estimate is found when the LENA system is applied to American English [10, 11]. Interestingly, this is also the study in which the human word count estimate was a bit on the "quick and dirty" side. Instead of listening carefully and transcribing everything they heard, listeners in this study just tap-counted the words they heard within short segments. This corresponds to yet unpublished data from Stockholm Babylab, in which the same procedure was followed as in the present study, except that the transcribers were restricted in the number of times they could replay the audio. Any words the transcribers were unable to discern during those two repetitions were not included in the transcription, and thus not in the word count. The purpose of this approach was to increase the ecological validity of the listening situation, as one cannot rewind real life. With this “quick and dirty” approach to transcriptions, the correlation between the human word count and the system’s AWC estimate is much higher than in the present data [17].

This highlights how the AWC estimate should be regarded: it is in fact an *estimate*, it cannot in any way or shape be used as an actual accurate count of the words spoken in the recording. The LENA system does not presently come close to matching the fine-grained perceptual skills of human listeners, especially not if listeners have the possibility to listen several times to, for example, sections of overlapping speech. But then again, as long as the estimate is regarded as an estimate, it can still be very useful, both in research and especially in clinical applications, as long as accuracy, precision and reliability are documented and taken into account when using it.

For this reason, evaluations of systems such as LENA are very important, in order to map out their scope and limitations, so that they can be used appropriately.

While the present study is a first step in an evaluation of the LENA system applied to Swedish, there are many steps left before the evaluation can be considered complete. There is a need for similar studies like the present one on the other two estimates delivered by the system, Conversational Turn Count (CTC) and Child Vocalization Count (CVC). These will have to include multiple ages of recorded children since both of those estimates are dependent of the reported age of the target child.

Further, it is also necessary to evaluate the automatic segmentation that is the basis of the estimates [18]. There have been reports of instances where the LENA system categorized large portions of the audio recording erroneously, for example speech in a TV program being segmented as male speech [19] or an elderly woman being categorized as a child [15]. Any automated segmentation based on acoustic characteristics of the signal is expected to make some erroneous predictions. However, more research is necessary in order to establish how common those instances are in the LENA system.

Despite its shortcomings, LENA is certainly a very useful tool to study child language environment at home on a large scale. The speed and ease of use makes the LENA system highly applicable for clinical interventions. What is needed though before starting out on a wide-scale use of LENA on various languages, is a proper evaluation for each of these languages.

5. Conclusions

The present study is the first step in evaluating the LENA system applied to Swedish. The reliability of the AWC as estimated by the LENA system when applied to Swedish was found to be within the range of its reliability for Spanish, Mandarin and French [2, 15, 16], and most similar to its reliability for French [16]. Further evaluations of other estimates and different aspects of the LENA system are crucial for any language to which it is applied, in order for it to be used appropriately in research and clinical applications.

6. Acknowledgements

The present study was funded by the Marcus and Amalia Wallenberg Foundation (MAW 2011.0070, MAW 2013.0056) and the Department of Linguistics, Stockholm University. The authors would also like to thank all families contributing with day-long recordings, and Elísabet Eir Cortes, Lena Renner and Johan Sjons for helpful comments on an earlier version of the manuscript. A big thank-you goes to our reviewers and to the participants of WEIRCLE and DARCLE for their inspiration.

7. References

[1] J. Huttenlocher, W. Haight, A.S. Bryk, M. Seltzer, and T. Lyons, "Early vocabulary growth: Relation to language input and gender". *Developmental Psychology*, vol. 27, no. 2, pp. 236-248, 1991.

[2] A. Weisleder and A. Fernald, "Talking to children matters - Early language experience strengthens processing and builds vocabulary". *Psychological Science*, vol. 24, no. 11, pp. 2134-2152, 2013.

[3] N. Hurtado, V.A. Marchman, and A. Fernald, "Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children". *Developmental Science*, vol. 11, no. 6, pp. F31-F39, 2008.

[4] E. Hoff and L.R. Naigles, "How children use input to acquire a lexicon". *Child Development*, vol. 73, no. 2, pp. 418-433, 2002.

[5] N. Hsu, P.A. Hadley, and M. Rispoli, "Diversity matters: Parent input predicts toddler verb production". *Journal of Child Language*, vol. 44, no. 1, pp. 63-86, 2017.

[6] F.J. Zimmerman et al., "Teaching by listening: The importance of adult-child conversations to language development". *Pediatrics*, vol. 124, no. 1, pp. 342-349, 2009.

[7] B. Hart and T.R. Risley, *Meaningful differences in the everyday experience of young American children*, Baltimore, MD: Brookes Publishing Company, 1995.

[8] B. Hart and T.R. Risley, "The early catastrophe: The 30 million word gap by age 3". *American Educator*, vol. 1, pp. 1-7, 2003.

[9] M. Ford, C.T. Baer, D. Xu, U. Yapanel and S. Gray, The LENATM Language Environment Analysis System: Audio specifications of the DLP-0121. LENA Foundation Technical Report LTR-032, Boulder, CO: LENA Research Foundation, 2008.

[10] J. Gilkerson and J.A. Richards, *The LENA natural language study. LENA Foundation Technical Report LTR-02-2,* Boulder, CO: LENA Research Foundation, 2008.

[11] D. Xu, U. Yapanel, and S. Gray, Reliability of the LENATM Language Environment Analysis System in young children's natural home environment. LENA Foundation Technical Report LTR-05-02, Boulder, CO: LENA Research Foundation, 2009.

[12] J.B. Oetting, L.R. Hartfield, and S.L. Pruitt, "Exploring LENA as a tool for researchers and clinicians". *ASHA Leader*, vol. 14, pp. 20-22, 2009.

[13] M. VanDam, and N.H. Silbert, "Precision and error of automatic speech recognition". *Proceedings of the Meetings of the Acoustical Society of America*, vol. 19, pp. 060006, 2013.

[14] M. VanDam and N.H. Silbert, "Characteristics of automatic and human speech recognition processes". *Journal of the Acoustical Society of America*, vol. 134, pp. 4096, 2013.

[15] J. Gilkerson et al., "Evaluating Language Environment Analysis system performance for Chinese: A pilot study in Shanghai". *Journal of Speech, Language and Hearing Research*, vol. 58, no. 2, pp. 445-452, 2015.

[16] M. Canault, M.-T. Le Normand, S. Foudil, N. Loundon and H. Thai-Van, "Reliability of the Language ENvironment Analysis system (LENATM) in European French". *Behavior Research Methods*, vol. 48, no. 3, pp. 1109-1124, 2016.

[17] I.-C. Schwarz, J. Schelhaas, and E. Marklund. (in preparation). Comparing different transcription strategies in child language acquisition research.

[18] M. Van Dam and N.H. Silbert, "Precision and error of automatic speech recognition", *Journal of the Acoustical Society of America*, vol. 133(5), pp. 3245, 2013.

[19] H. Elo, "Preliminary observations of feasibility of LENA with Finnish". Paper presented at the 1st Meeting for LENA users in the Nordic countries. Stockholm, Sweden, 2016.