1   **Temporal Reliability of 10 Language Measures in Adults: Reliability of Morpheme**

2   **Production in Adult Speech Compared to Children's**

3   **Abstract**

4   Language samples are used for several purposes in both healthy and clinical populations,

5   but little information exists about the length a language sample needs to be to accurately

6   represent a person's linguistic skills, especially for adults. This study used a test-retest paradigm

7   to compare language samples taken from two conversations of 20 adults speaking in pairs,

8   recorded a maximum of a week apart. Frequency counts and reliability measures of several

9   inflectional morphemes were compared from recording 1 to recording 2 for sample sizes of 50,

10   100, 150 and 200 utterances. Results showed sizable variation in frequency and reliability

11   between different morphemes, suggesting that some linguistic items may be more useful

12   measures of typical adult language than others. 100 utterances appeared to be the most useful

13   sample size, after taking into consideration the degree of reliability and time constraints of

14   gathering samples. Finally, a comparison of adult data to earlier child reliability data suggested

15   that if reliability is used as a benchmark of language proficiency, higher reliability is likely to be

16   a sign of greater proficiency.

17   **Keywords**

19   Linguistic output can be measured and analyzed in numerous ways, dependent upon the

20   goals of the analyst. Whereas standardized language tests typically limit themselves to the

21   domains for which they were designed, samples of spontaneous language can be used to

22   investigate an extensive range of language related areas such as word-finding (German 1987),

23   vocabulary (Bleses, Werner & Philip 2018), grammar (Hollister, Van Horne & Zebrowski 2017),

24   pragmatics (Klusek, Martin & Losh 2014), and language dominance (Solorio et al. 2011).

25   Language samples can also be put to clinical use in areas as diverse as exploring grammatical

26   ability in children with autism spectrum disorder (Wittke et al. 2017), comparing the language of

27   groups with Parkinson's disease differentiated by neural degeneration in the right versus left

28   hemisphere, (Batens et al. 2015), tracking language development in young children with hearing

29   loss (Tomblin et al. 2015) and even the analysis of a presidential Twitter feed (Ott 2017).

30   The value of spontaneous speech samples was made clear by Wagenaar, Snow and Prins

31   (1975) who stated that they "… provide[s] the most subtle and complete reflection of language

32   abilities". However, this leads to the recognition of a serious problem; despite the widespread

33   use of language samples in research, the issue of reliability, or in other words, the issue of how

34   representative any given sample is of an individual's language, is mostly unknown. For example,

35   if a person's speech is recorded until a specified amount of utterances have been produced, it is

36   unclear at what point we can say that the sample is long enough to be representative of that

37   individual's typically produced language. Furthermore, when considering the notion of being

38   representative of a person's linguistic output, a nuanced question must specify a precise area of

39   language, as it is unlikely that the same size sample would be the required amount for an honest

40   representation of phonological, semantic and morphosyntactic capacities. Instead, each area will

41   be found to have its own required sample size.

42   In addition to the question of the size of the language sample necessary to provide an

43   accurate representation of a person's typical speech, further issues remain unexplored such as

44   whether it makes a difference if the sample belongs to an adult versus a 3-year-old child. Our

45   inability to answer these questions was emphasized by Muma (1998: 316) who warned that the

46   often-used sample size of 50-100 utterances in the clinical study of child language is in no way

47   guaranteed to be an accurate representation of their typical language. The issue is further

48   highlighted by Tomasello and Stahl (2004) who warn researchers of the potential risk of error

49   when assuming that a sample is representative of a speaker's ability as any sample is likely to

50   represent only a small portion of a given person's maximum linguistic ability. Cole et al. (1989:

51   260) emphasized that "although reliability information is basic to the interpretation of test

52   results, this measurement characteristic appears to have been generally overlooked in the area of

53   language sample interpretation". Despite Cole's warning over 30 years ago, little work has been

54   published to remedy this issue, particularly around adult language.

55   The goal of this study is to investigate the reliability of language production in healthy

56   monolingual English-speaking adults engaged in spontaneous conversation with other adults.

57   Degrees of reliability will be established for samples of 50, 100, 150 and 200 utterances to

58   determine whether increasing sample sizes correlate with increasing reliability. This information

59   will in turn allow the authors to suggest which morphemes are more reliable than others which

60   may be useful for future clinical studies. It will also lead to a discussion about what sample size

61   out of those tested is the most useful for research. The specific items examined will be a group of

62   inflectional morphemes used in child studies carried out by Author (2013; 2014; 2015). Finally,

63   adult frequency and reliability will be compared to help determine whether reliability increases

64   with greater mastery over language or whether it decreases over time, due to using a greater

65   variety of linguistic forms.

66   Different types of reliability exist and more than one can be appropriately used with

67   language samples, but the type discussed in this article is known as test-retest reliability or

68   temporal reliability. This type of reliability, which will be referred to throughout the remainder

69   of this article simply as 'reliability' measures the stability of the scores from the same person at

70   two or more different times using the same measurement tool. For example, tests of IQ are

71   carefully tested for temporal reliability to ensure that the test will provide a similar score for an

72   individual each time taken. In this study, temporal reliability is used to assess to what degree the

73   production frequency of a particular item derived from a language sample is consistent with the

74   production frequency of the same item taken from a language sample under similar

75   circumstances but at a different time. For an extensive discussion of the statistical concept of

76   reliability under Classical Test Theory (CTT), refer to Author (2014).

77   **Motivations for the Current Study**

78   Numerous motivations underlie this study; the first being a desire to understand the

79   frequency and reliability of specific morphemes in adult language. Although a small number of

80   studies, which will be discussed shortly, exist regarding reliability of general measures of adult

81   language such as mean length of utterance (MLU), little is known about the reliability of

82   syntactic constructions as specific as individual bound morphemes. Having such limited

83   knowledge regarding the reliability of typical adult human language, at least in the area of

84   specific syntactic structures such as inflectional morphemes, leads to the question of whether

85   reliability, or frequency for that matter, are possible candidates for providing a measurable

86   baseline or benchmark representing typical native adult levels of language production, especially

87   if those baselines differ from children's.

88   Another motivation is to continue to explore the relationship between frequency and

89   reliability. Work by Author (2013; 2014; 2015) indicated a general higher level of reliability in

90   morphemes produced more frequently, including the highest reliability of all tested items for a

91   specially developed target called 'multiverb' which was a global measure counting any utterance

92   with more than one verb. This item was included specifically to ensure that 2 and 3-year-old

93   children had at least one item with high frequency so this link could be explored. For individual

94   syntactic structures, the copula had the highest frequency and highest reliability at each sample

95   size while the genitive was the least frequent and displayed the lowest degree of reliability.

96   Interestingly, the copula stood out from the other morphemes tested when a split-half comparison

97   was carried out, showing that while most morphemes tended to be spread relatively evenly

98   throughout the samples, use of the copula was more clustered.

99   Another important motivation is to ensure that professionals who use language samples

100   are aware of the issues around minimum sample sizes required to be representative of a person's

101   linguistic skills. Without the knowledge of the reliability of most aspects of language,

102   conclusions drawn from language samples may be considered suspect. For instance, language

103   programs often offer a test of English for international students when they first arrive to the US.

104   Along with a multiple-choice placement test, students generally produce a single writing sample

105   and participate in a brief oral interview. On the basis of this testing, students are either placed

106   into a particular level or released from the requirement to attend English language classes before

107   matriculating to credit courses. In other words, these are high stakes tests, but they rely on

108   limited language samples. To our knowledge, no research has been done on the required sample

109   size of oral or written samples of a foreign language to show the appropriate minimum sample

110   size.

111   These same limitations also apply to professionals diagnosing and treating individuals

112   with language-related impairments such as Specific Language Impairment and aphasia. It is

113   crucial that the degree of reliability of the sample be considered, given that language samples are

114   commonly used as an important aspect of diagnostics and measurement of progress (Simon-

115 Cereijido & Gutiérrez-Clellen 2007; Zimmerer et al. 2020) In order to do this though, we must

116 first examine the degree to which language samples taken from healthy adult populations can be

117 considered reliable.

118 **Review of the Literature**

119   A review of the literature shows that much of the reliability work carried out in adult

120 language samples has been in the field of aphasia, language disability caused by brain trauma,

121 most often post-stroke. Unfortunately, the body of work examining reliability of both healthy

122 adults and aphasia patients was small enough to be referred to as "relatively unexplored

123 territory" by Prins and Bastiaanse in 2004. That categorization is still applicable today.

124 Aphasia can be divided into fluent and non-fluent categories and one characteristic of non-fluent

125 aphasia is agrammatism, meaning that speech is made up primarily of content words and lacking

126 in function words. A study by Saffran, Berndt and Schwartz (1989) examined the reliability of

127 language in patients with agrammatic aphasia, looking at the proportion of closed class words,

128 proportion of verb inflections, proportion of well-formed sentences and an embedding index.

129 Measures of temporal reliability varied from .53 to .92, leading Prins and Bastiaanse (2004) to

130 remark that "It may be the case, however, that this variability is not due to unreliability of the

131 scoring system, but to the unstable behavior of the agrammatic speakers", specifically referring

132 to a lack of free and bound morphemes and a lack of well-formed sentences as hallmarks of

133 agrammatic aphasia. Variability in language production in aphasic speakers underscores the

134 importance of understanding reliability of spontaneous language in clinical groups.

135   In 1993, Nicholas and Brookshire carried out reliability checks on their newly developed

136 standardized, rule-based language scoring system, quantifying speech informativeness of

137 aphasics based on the analysis of Correct Information Units (CIUs). CIUs are topics pre-

138 identified as a response to specific pictures used as tools to elicit speech. In addition to the

139 number of CIUs per minute, words per minute (WPM), percentage CIUs and total number of

140 words were monitored. Four individual pictures, two picture sequences and two direct questions

141 were used to elicit speech samples. The researchers analyzed the performance of 20 aphasics

142 and 20 healthy adults to explore group differences including reliability. Each individual was

143 tested three times in order to determine the reliability of production of each linguistic item.

144 Correlations in both healthy and brain-damaged adults ranged from .88 to .98, showing high

145 reliability for both groups, although WPM, CIUs per minute and percentage CIUs were more

146 stable than the number of CIUs and number of words. Notably, the lowest reliability for any

147 measure was r = .88 for non-brain-damaged people for the measure of CIUs per minute. A

148 subsequent paper by the same authors (Brookshire & Nicholas 1994) reexamined the data from

149 their 1993 study, analyzing language samples derived from a smaller subset of the 10 original

150 stimuli used in the original paper. Results showed that reliability tended to drop as the number of

151 stimuli declined and that when using one picture as the sole stimulus, some subjects exhibited

152 "dramatic instability" of language production. The authors warned of making decisions about the

153 speech of aphasics based on one short sample, "because such measures can be highly unstable

154 from test to test. Because of this instability, a patient's type or severity of aphasia might appear

155 to have changed, even though no actual change has occurred."

156   The Nicholas and Brookshire (1993) tool was criticized by Armstrong (2000) speaking of

157 the "lack of adequate linguistic description" as well as the need to use more stimuli in order to

158 increase the length of the language samples in order to be representative of the speakers' overall

159 abilities.

160 Given the possible propensity to greater variation in speech production of patients with

161 aphasia than healthy adults, it is important to have reliability measures of these healthy adults for

162 the sake of comparing aphasic and typical language production. This is highlighted by Prins and

163 Bastiaanse (2004) who made the following recommendation regarding research into language

164 sample use in aphasia: "One of the topics that should be investigated is the reliability of

165 linguistic variables. Although inter-rater reliability is usually satisfactory, hardly anything is

166 known about test-retest reliability. In other words: to what extent is the linguistic behaviour of

167 aphasic patients stable? Furthermore, especially for clinical practice, it is important that group

168 studies are performed in order to establish standards for statistically reliable improvement. In this

169 way, the results of individual treatment studies could be interpreted much better."

170 To develop performance norms in adults of different ages, a study by Wright et al. (2005)

171 looked specifically at reliability measures in 40 healthy adults who were divided by age into a

172 younger group (mean 23.9 years) and an older group (mean 67.6 years). Language samples were

173 gathered on two occasions, 10-20 days apart for to examine test-retest reliability scores of 'main

174 events', which are identical to CIUs discussed earlier. Language samples were gathered based on

175 elicited responses to two single pictures and two picture sequences taken from Nicholas and

176 Brookshire (1993). If a response lasted for a duration of less than 15 seconds, participants were

177 encouraged to speak more. While the younger group displayed a significantly higher number of

178 main events (or CIUs) than the older group, the reliability of the younger adults' main event

179 production was .70 while the older group's was .76, a result they called "sufficiently stable".

180 Unfortunately, the authors did not report the length of the language samples used although they

181 do highlight the need to have norms for a non-brain-damaged population that can be used for

182 age-matching in further research.

183   When exploring the body of work that has been carried out in the field of reliability of

184   adult language production, it is notable that several different target items have been examined

185   such as words produced per minute (Saffran, Berndt & Schwartz 1989; Nicholas & Brookshire

186   1993), number and percentage of CIUs (Nicholas & Brookshire 1993), proportion of well-

187   formed sentences (Saffran, Berndt & Schwartz 1989), degree of embedding (Saffran, Berndt &

188   Schwartz 1989) and inclusion of main events (Wright et al. 2005). These measures could be

189   referred to as *global*, meaning that they are not looking at the use of particular words or syntactic

190   constructions. The literature is lacking both frequency and reliability information regarding

191   specific linguistic items in adult language. However, a small number of studies of child language

192   have taken a more microscopic view of language, examining individual morphemes which could

193   provide a foundation for adult work.

194   The first and largest of these studies to date is from a dataset published by the Wisconsin

195   Department of Public Instruction (Leadholm & Miller 1992), which collected language samples

196   from 266 children in local schools, ranging from 3 – 13 years of age. The goal of the study was

197   to develop norms for a typical population in order to provide a comparison for language

198   produced by children with language disability but only frequency**,** range and standard deviation

199   were provided without reliability measures. Each age has data for several global items provided

200   for both 100 utterance samples and 12-minute samples including MLU, total number of words

201   and number of different words. What sets this study apart however, is the inclusion of frequency

202   information of specific syntactic constructions including the bound morphemes of regular past,

203   plural, possessive, third person singular and present progressive along with personal pronouns,

204   conjunctions and modals. Semantic counts are also available for individual question words,

205 conjunctions, negatives, modals, and pronouns, providing the most extensive frequency norms of

206 specific items in child language to date.

207   Two studies based on 27 children aged 2;6 to 3;6 (Author 2013; 2014) looked at the

208 frequency of production of specific morphological constructions in the same vein as Leadholm

209 and Miller (1992), but also examined reliability of different sample sizes using a test-retest

210 procedure based on counts of the copula, past tense, third person singular, *do* as an uncontracted

211 auxiliary, *be* as an uncontracted auxiliary, the contracted auxiliary, the genitive and a one more

212 global structure that they termed 'multiverb utterances' which was defined as being any

213 utterance with more than one verb. The target morphemes were chosen based on the theory of

214 specific disruption to certain inflectional morphemes in children with Specific Language

215 Impairment (Leonard, Eyer, Bedore & Grela 1997), with the goal being to provide a baseline of

216 typically developing children that might be of assistance in developing tools to recognize

217 children with SLI at an early age. Frequency counts of each item were provided for sample

218 lengths of 50, 100, 150 and 200 utterances to compare reliability. Results ranged from -0.01 for

219 the past tense at 100 utterances to 0.78 for contracted auxiliaries at 200 utterances, showing that

220 the shift to counting specific linguistic structures led to largely varying and overall lower

221 reliabilities than the global measures examined earlier.

222   Author (2015) next examined the differences between child and adult speech in terms of

223 the frequency and reliability on the same morphemes as their earlier studies, again using sample

224 sizes of 50, 100, 150 and 200 utterances taken from 17 mothers interacting with their children

225 aged 2;6-3;6, thereby creating samples of child-directed-speech, known for its difference from

226 adult to adult speech due to its qualities of being "syntactically and semantically simpler than

227 [language] used to address adults" (Grieser and Kuhl 1988). Reliabilities in child-directed adult

228   language varied from -0.05 for the third person singular at 50 utterances to .58 for the plural at

229   200 utterances. Results indicated that frequency counts were higher for adults than the children

230   aged 2;6 – 3;6, the relative frequency of items in the child language samples mirrored those in

231   adult samples and that the reliability of most elements examined was lower in adult language

232   than child language. Although of interest, the fact that the adults were speaking to their very

233   young children means that it is likely that their speech in this context was not representative of

234   their typical language with other adults.

235   In addition to developing frequency and reliability norms of healthy adults producing

236   conversational language, this study also compares the speech of young children to the speech of

237   adult-to-adult language, to determine whether significantly different production patterns may

238   provide markers of language proficiency. If it is discovered that young children's reliability

239   varies greatly from adult language in terms of reliability, this could suggest that reliability is a

240   measure of competency. There are reasons to think that this could work in different directions. It

241   is possible that higher reliability is a hallmark of early language learning, in that fewer words and

242   structures are available to the child (or foreign language learner), therefore creating a situation

243   where the same structures are used more repetitively in early language, leading to higher

244   reliability. On the other hand, it is possible that adult language will be more reliable than

245   children's given the length of utterances which leads to higher frequency of usage of the

246   examined morphemes. Either way, a marked discrepancy between the two groups could suggest

247   that the reliability measure assigned to adult language is a 'goalpost' towards which early

248   language would move as it develops.

249   This study seeks to answer the following, specific research questions:

250   a. How frequent is the production of each of a defined set of morphemes in adult-to-

251   adult speech at 50, 100, 150 and 200 utterances?

252   b. How reliable is the production of these morphemes at sample lengths of 50, 100,

253   150 and 200 utterances?

254   c. How do frequency and reliability measures of specific bound morphemes and the

255   more general multiverb in adult language samples compare to those in Author's

256   (2013; 2014) child language samples?

257   **Method**

258 **Participants**

259   Transcripts from recordings of 20 adults (13 female, 7 male) aged 18-65 (mean 33.4,

260 SD=15.6) were analyzed in this study. All adults recruited were native English speakers living in

261 monolingual homes and reported negatively to having hearing problems, current or past speech

262 or language disorders, speech and language therapy or neurological issues that might have an

263 effect on language, speech or cognitive ability. All participants also responded that there were no

264 immediate family members who had been diagnosed with speech or language issues.

265 **Procedure**

266   Participants in this study were recorded in two sessions within a week of each other in the

267 location of their choice based on instructions of being in a quiet place where they would not be

268 interrupted and that the location be the same each time. Recordings were made on smartphones

269 and participants were asked to talk to each other about topics of their choosing for approximately

270 30 minutes each session. Although instructions to speak about past events were included in

271 Author's earlier studies of children aged 2;6-3;6 in order to offer the opportunity to produce the

272   past tense, it was deemed unnecessary in this case due to the volume of language expected to be

273   produced and the high degree of speaker competency which was viewed as likely to produce a

274   variety of tenses naturally.

275   Orthographic transcription of the samples was carried out by university graduate students

276   in the first author's class as an exercise. These transcriptions were then reviewed and corrected

277   by the second author. Each sample was divided into utterances according to P-units (Loban

278   1976). Following Miller and Chapman (2004), P-units were limited to a maximum of two

279   independent clauses in order to avoid run-on sentences. 5% of samples transcribed by the second

280   author were transcribed by the first author to assess transcription reliability. Inter-transcriber

281   reliability was 0.94.

282   Identical rules about inclusion and exclusion from the sample from Author (2013; 2014)

283   were applied to the current study. For example, a word that was repeated due to lack of fluency

284   was counted only once; if a participant uttered "I kicked… kicked the ball", only one count of a

285   past tense would be awarded for the utterance. It also would not be scored as a multi-verb

286   utterance which is described below. Each item or category had a list of examples of inclusion vs

287   exclusion of questionable occurrences. One example of this is shown by the following utterance

288   when considering whether or not a copula had been used:

289   "I was like 'come on over'"

290   Due to the repetitive nature of this structure, in which 'was' combined with 'like' meant

291   'said', it was decided that this particular form of 'was' would not be included as a copula. A

292   similar decision was made not to include words such as 'clothes' into the category of plural as it

293   does not show productive plurality of a noun. More important than exactly what was included

294   and excluded from a morphosyntactic category such as 'genitive' is the fact that the same rules

295   were applied consistently across all participants and all studies to ensure true comparisons of

296   child and adult reliability.

297   The items chosen to examine in this study were based on the earlier child studies by

298   Author, (2013; 2014). The former of these papers focused on morphosyntactic items that have

299   been put forward as potential markers for Specific Language Impairment (Cleave & Rice 1997;

300   Leonard 2014; Leonard et al. 1997; Rice et al. 1995; Rice & Wexler 1996) including the copula,

301   third person singular (3s), *do* as an uncontracted auxiliary (*do*-aux), *be* as an uncontracted

302   auxiliary (*be*-aux), contracted auxiliary ('-aux), and past tense (-ed). The latter paper added items

303   including the plural and the present progressive (-ing) from Stage II of Brown's table of

304   morpheme acquisition (1973) and the genitive (-'s possessive) from Stage III in order to compare

305   the reliability of more recent Stage III acquisitions with those of the earlier acquired Stage II.

306   Both papers also included a newly constructed category labelled as "multi-verb utterance",

307   defined as any utterance with more than one verb, including auxiliaries with a lexical verb or

308   verbs in separate clauses. The sole purpose of the multi-verb was to ensure that there was an item

309   that was likely to have high frequency even in the language produced by children at age two in

310   order to allow the researchers to ask the question of whether high frequency items are more

311   reliable than less frequent items.

312   [TABLE 1 NEAR HERE]

313   Multi-verb was the only binary item from the list. Whether an utterance contained two or

314   more verbs, it got a single count of 1 as opposed to 0 which was awarded to any utterance free of

315   verbs. All other items received a count for the number of times they appeared in the sample with

316   no ceiling.

317   Each conversation's transcript was a minimum of 225 utterances. The first 25 were

318   excluded for the purpose of allowing a brief warmup phase and to match the procedure used in

319   Author's earlier studies (2013; 2014). The ensuing 200 utterances were coded according to the

320   number of occurrences in each utterance of each targeted morphosyntactic item. Overall totals

321   for each target item were established for each item at 50, 100, 150 and 200 utterances.

322   **Results**

323   **Adult Frequency**

324   In the following table, each morphosyntactic item is listed twice and identified with a 1 or

325   2. This indicates whether it refers to the first or second language sample gathered.

326   [TABLE 2 NEAR HERE]

327   In looking only at frequency of adult language in Table 2, it is clear that some structures

328   are much more common in adult speech than others. For instance, at 200 utterances, uses of

329   multiverb constructions, copula, and past tense number close to or above 50 instances. On the

330   other hand, genitive and 3rd person singular still number fewer than 10, even at 200 utterances.

331   Plural, -ing, *be*-aux, *do*-aux, and 'aux all have between 13 and 30 instances of use. From this we

332   can confirm that different morphosyntactic structures are not used at equal frequencies, even in

333   adults with fully formed linguistic systems. Clearly, some structures are much more frequent

334   than others, as can be seen in Table 3 In order from most frequent to least frequent, based on

335   counts at 200 utterances with the right column showing their average frequency in a 200-

336   utterance sample:

337   [TABLE 3 NEAR HERE]

338   [TABLE 4 NEAR HERE]

339   When we look at the correlations for adults in Table 4, we see that there are several

340   robust correlations. At 50 utterances, only multiverb and copula reach a significant correlation,

341   but by 100 utterances, plural, 'aux, and past have all reached significant correlations that

342   continue through 150 and 200 utterances. The correlations for both -ing and be-aux reach

343   significance at 100 utterances, but drop to .5 or below at 150 utterances. Genitive, do-aux, and 3s

344   simply never reach a significant correlation.

345   This data speaks to our question of what morphosyntactic structures are appropriate for

346   study with language samples. First, several of these structures, including genitive, do-aux, and

347   3s, do not appear to be reliable. They never reach significant correlations, even at 200 utterances.

348   For the purposes of language sample analysis that relies on consistent occurrences of a structure,

349   none of these three are reliable enough for dependable analysis. In other words, these

350   constructions do not appear to be used consistently by adults in different language samples, and

351   are unlikely to be useful in assessing an individual's language proficiency unless much longer or

352   elicited samples in future studies show much higher reliability.

353   At the other end of the spectrum are the structures with robust correlations, indicating

354   that these structures are produced consistently across language samples. These are multiverb,

355   copula, plural, and 'aux, all of which reach correlations of .7 or above in samples of 100

356   utterances or fewer. Of these, multiverb and 'aux have the strongest correlations, giving them the

357   greatest potential for providing benchmarks in language sample analysis.

358   Between these two we have a small set of structures that are inconsistent, or for which

359   correlations are not robust. Following Wright et al. 2005, we consider .7 to be sufficiently stable

360   for our correlations, and none of these forms reach a correlation of .7, though they do reach

361   correlations of .5 or above. These structures may show themselves to be of use in language

362   proficiency measurement, but perhaps not to the same degree as the structures and measures

363   already shown to be more reliable. This medium category includes -ing and be-aux, both of

364   which reach a correlation between .56 and .61 at 100 utterances, but then drop to .5 or below at

365   150. Here we also include past, which reaches a correlation of .6 by 50 utterances, but this

366   correlation never rises to .7.

367   [TABLE 5 NEAR HERE]

368   In addition to illustrating the reliability of some particular measures, we must also

369   consider the sample size for these claims. It is clear that for most morphosyntactic measures

370   under examination for adults, reliability is not present at 50 utterances. However, for most

371   selected morphosyntactic measures, a significant correlation is reached by 100 utterances. At that

372   point, only genitive, do-aux, and 3s fail to reach a significant correlation. Furthermore,

373   correlations for these three morphemes never rise above .47, and that is only for genitive; for do-

374   aux, .1 is the strongest correlation, while the correlation for 3s is always negative. In short, the

375   items that have not reached a significant correlation by 100 utterances do not reach one by 200.

376   While the correlations for multiverb, copula, and 'aux continue to strengthen with longer

377   samples, the gains are all .11 or less, and in others, like plural, reliability falls slightly with

378   longer samples. This data seems to indicate that 100 utterances is the optimal sample size for

379   reliability of those tested here; at 150 utterances, -ing and *be*-aux both fall to .5 or below, and

380   while -ing returns to a significant correlation at 200 utterances, *be*-aux does not. This instability

381   in the strength of the correlation is seen in several other instances as well: at 100 utterances the

382   plural has a correlation of .8, but this correlation has fallen to .69 at 200 utterances. The data thus

383   indicates that 100 utterances should be enough for a reliable sample; it is not clear that the gain

384   in strength of correlation is enough to merit the time and effort involved in collecting and

385   transcribing an additional 100 utterances. Thus, if we are looking for how long a language

386   sample needs to be in order to be considered a reliable sample of a person's use of inflectional

387   morphemes, a sample of at least 100 utterances is recommended to ensure that studies of adult

388   language are scientifically meaningful.

389   [TABLE 6 NEAR HERE]

390   [TABLE 7 NEAR HERE]

391   **Comparison of Adult and Child Reliability**

392   When we compare adult and child language samples, there are a number of similarities.

393   First, the child language samples reported in Author (2013; 2014) reflected varying frequencies

394   for different morphosyntactic forms. Multiverb was the most frequent, at 37 and 44 uses at 200

395   utterances, while genitive was the least frequent with values close to 1 even at 200 utterances.

396   For children, this variation could have been at least partially due to children still being in the

397   process of acquiring this structure, but this explanation is not available for the adults, whose

398   linguistic systems are assumed to be fully formed. For both children and adults, we see that

399   genitive forms are only rarely used, even in mature linguistic systems.

400   When we compare adult and child language, we further see that the relative frequency is

401   not identical. Table 8 shows the relative frequency of adult and children at 200 utterances (most

402   frequent of the 2 samples taken).

403   [TABLE 8 NEAR HERE]

404   In both cases, multiverb is the most frequent and genitive the least frequent. However, the

405   other structures vary in relative frequency between the 2 groups. For instance, the frequency of

406   past tense is second only to multiverb for adults, with an average of 64.2 uses, while for children,

407   the past is only used an average of 4.5 times, falling close to the middle in terms of relative

408   frequency for the group as a whole. This difference in relative frequency could be due to several

409   factors. First, the children in the study were ages 2;6 – 3;6 and were potentially still in the

410   process of acquiring some parts of their linguistic system. Therefore, they may not have fully

411   acquired the past tense. Adults, on the other hand, have a fully developed linguistic system, and

412   could have a stronger command of past tense usage. In addition, it is likely the case that adults

413   speaking to other adults tend to talk more, although they may engage more in discussion of the

414   here and now with children. This leads to the question of whether adult language varies

415   morphosyntactically depending whether they are speaking to a child or another adult. Although

416   of interest, this question is beyond the scope of the current work.

417   Additional comparisons with the correlations from Author's child language study (2015),

418   shows some striking similarities between the child and adult language samples, as well as several

419   distinct differences. First, multiverb is the strongest correlation at 50 utterances in both adults

420   and children. As can be seen in Table 8, this is also the most frequent structure. Second, the most

421   robust correlations arise with multiverb, copula, and 'aux, but the correlations are stronger in the

422   adult language. For adults, all three of these reach correlations above .8, with 'aux as high as .89.

423   For children, these same measures fall between .63 (copula) and .78.

424   Another similarity evident between the adult and child correlation patterns is fluctuation

425   within the strength of correlation (previously discussed for adults). For example, in the child

426   samples, correlations for the copula reach .7 at 150 utterances, but fall to .63 at 200 utterances.

427   This fluctuation leads to the conclusion that these correlations are driven by more than just

428   frequency. Because each sample is nested within those with greater utterances, clearly a sample

429 of 200 utterances will have an equal or greater frequency across all categories than a sample of

430 100 or 150 utterances. If reliability was solely driven by frequency, we would expect to see

431 correlations rise steadily from 50 to 200 utterances. However, this is not what is seen in Tables Y

432 and Z.

433   These fluctuations speak directly to the question of how long a language sample should

434 be in order to be considered reliable. Based on the low correlations present between 50-utterance

435 samples, Author (2015) claim that 50 utterances for a child language sample is not reliable

436 enough for the morphsyntactic measures at hand. Here, we see that the same claim must be made

437 for adult language samples: 50 utterances is not enough to see reliable use of morphosyntactic

438 structures. While previous researchers (Pavelko, Price & Owens 2020) have found reliability in

439 samples even as short as 25 utterances, the measures under examination in those studies were

440 more global measures, like MLU, words per sentence, and clauses per sentence. For the usage of

441 specific inflectional morphemes, correlations at 50 utterances are just not strong enough to draw

442 conclusions about the use of these items. A sample of at least 100 utterances is necessary.

443   The most robust correlations for adults, falling at .7 or above at 100, 150, and 200

444 utterances, are multiverb, copula, and 'aux. These three constructions are clearly the most

445 reliable. However, when comparing these with the frequencies previously shown, they are not all

446 the most frequent. While multiverb and copula are both quite frequent, 'aux is much less

447 frequent, appearing only an average of 13.55 times even in a 200-utterance sample. Past, which

448 is the second most frequent construction, does have significant correlations, but these never rise

449 to .7, even though frequency rises to an average of 76 instances at 200 utterances in Time 2.

450 From this, we can see that frequency is not a direct indicator of the strength of correlation. In

451 other words, it is not the case that the strongest correlations always arise from the most frequent

452 items.

453   In their study of child language, Author (2013) found similar results for multiverb,

454 copula, and 'aux (see Table 7). Results showed that the strongest correlations found were for

455 multiverb, copula, and 'aux, which all reached correlations of .57 or above by 100 utterances,

456 and by 150 utterances had all reached .7. However, uses of plural never reached a significant

457 correlation at 100 and 150. On the other hand, *be*-aux was quite robust in the child language

458 samples, reaching a correlation of .74 by 100 utterances, but *be*-aux was not as robust in the adult

459 language; it reached a significant correlation of .61 at 100 utterances, but fell to around .5 at 150

460 and 200 utterances. Crucially, while multiverb and copula were among the most frequent items,

461 'aux and *be*-aux were much less frequent. Whether we are looking at children or adults, it is yet

462 again made clear that there is more than frequency driving the strength of correlation.

463   As we turn from frequency to examine other factors that may be affecting reliability, one

464 possibility to consider is that the most reliable forms of multiverb and 'aux are structures that do

465 not rely on any particular tense or content, unlike genitive, 3s, and *do*-aux, none of which reach

466 reliable correlations. In other words, use of the most reliable structures could be largely stylistic,

467 in the sense that they are not required for effective communication, but are used frequently

468 nonetheless. Given the option to use these forms, speakers use them with similar frequencies in

469 different samples. The less reliable structures, on the other hand, require some particular person

470 or situation in order to be used. The genitive requires speaking of possession, 3s requires

471 speaking of a person not involved in the conversation in present tense, and *do*-aux is most often

472 used in the asking of questions. In different conversations, these items may or may not be

473 appropriate given the conversation at hand. To ensure that these particular contexts exist in

474 different samples, it would be necessary to provide elicitation prompts or scenarios, but even this

475 would not guarantee a higher degree of reliability. No elicitation prompts were used for the adult

476 language samples, but in the child samples, parents were asked to talk about something that had

477 happened in the recent past. Even though adults complied, this did not result in strong

478 correlations for past tense in the child language samples.

479 The other two highest correlations for the adult language, plural and copula, do require

480 particular contexts for their use, but the contexts are not as specific as for genitive, do-aux, and

481 questions. In addition, the plural and copula structures are among the highest in frequency for

482 both children and adults. It is possible that these 2 forms are ones that appear frequently in

483 language in general, regardless of context. As we move toward further examination of reliability

484 in language samples, it may be fruitful to examine the degree to which additional morphemes

485 require particular contexts or participants.

486 Finally, we consider whether reliability measures may be a useful tool in measuring

487 linguistic competency where typical adult language of a native speaker is considered as the ideal,

488 and if so, is higher competency reflected by higher or lower reliability? Generally, samples of

489 child language were not as robust in either frequency or reliability as adult samples were, with

490 child reliability being lower than adult reliability in 28 out of the 40 spots on the reliability

491 charts. When considering reliability at only the 100 utterance measure, we find that for 7 out of

492 10 items, child reliability was lower than adult. This provides us with an indication that if

493 measures of reliability do indeed turn out to be a measure of linguistic ability, it seems likely that

494 lower reliability will indicate lower proficiency.

495 Because children were aged 2;6 – 3;6 and were still in the process of acquiring their

496 linguistic system, and Author's (2014) chosen morphosyntactic categories were chosen at least

497 partially due to their presence on Brown's order of acquisition list, one might think that low

498 reliability was dependent upon a lack of familiarity with specific target items counted. However,

499 the results overall cannot simply be shrugged off as due to acquisitional stages. Consider, for

500 instance, that contracted auxiliary is the latest acquired morpheme that Brown studied, later than

501 genitive, past, plural, 3s, all of which were less reliable in the child language study than 'aux

502 was. That the latest acquired morpheme was highly reliable is more likely to be related to the

503 nature of that morpheme than strictly about the acquisitional stages. In other words, there is

504 something about particular morphemes that makes them more likely to be used across a variety

505 of contexts and conversations, while other forms are less likely to be used in this way. This

506 finding is encouraging in that it suggests that certain reliable linguistic markers of development

507 may exist.

Conclusion

509   This examination of the reliability of adult language samples in comparison with those of

510 child language samples has shown several interesting results. First, we see that for both groups, a

511 sample size of 100 utterances is likely to be the best among those tested here, being long enough

512 to reach reliable correlations, but short enough for feasible analysis by linguists and clinicians.

513 Second, we see that frequency and reliability do not always go hand in hand, even in mature

514 linguistic systems given that the most robust correlations do not always arise from the most

515 frequent structures. Third, we have proposed that some morphosyntactic categories are better

516 candidates than others for providing linguistic benchmarks. In particular, the multiverb

517 construction, copula, plural, and contracted auxiliary all show robust correlations at 100

518 utterances; these structures are the best targets for morphosyntactic study that rely on quantity of

519 production. Additional forms, namely -ing, *be*-aux, and past, may be considered as reliable, but

520   to a lesser degree. Also, some morphosyntactic forms do not lend themselves to study due to

521   very low correlations. These forms include genitive, do-aux, and 3s. Finally, examination of the

522   reliability of adult morphosyntactic use in language samples provides early indicators that if

523   reliability will eventually be used as a measure of progress, it is likely that higher reliability, not

524   lower, will be a sign of greater proficiency.

525   **References**

526   Armstrong, E. 2000. 'Aphasic discourse analysis: The story so far,' *Aphasiology* 14/9: 875-892.

527   https://doi.org/10.1080/02687030050127685

528   Author 2015. Details withheld to preserve blind review.

529   Author 2014. Details withheld to preserve blind review.

530   Author 2013. Details withheld to preserve blind review.

531   Batens, K., De Letter, M., Raedt, R., Duyck, W., Vanhoutte, S., Van Roost, D., & Santens, P.

532   2015. 'Subthalamic nucleus stimulation and spontaneous language production in

533   Parkinson's disease: A double laterality problem,' *Brain and Language* 147: 76-84.

534   https://doi.org/10.1016/j.bandl.2015.06.002

535   Bleses, D., Werner, V. A. C. H., & Philip, S. D. 2018. 'Self-reported parental vocabulary input

536   frequency for young children,' *Journal of Child Language* 45/5: 1073-1090.

537   https://doi.org/10.1017/s0305000918000089

538   Brookshire, R. H., & Nicholas, L. E. 1994. 'Test-retest stability of measures of connected speech

539   in aphasia,' *Clinical Aphasiology* 22: 119-133. http://aphasiology.pitt.edu/163/1/22-

540   09.pdf

541   Cleave, P. L., & Rice, M. L. 1997. 'An examination of the morpheme BE in children with

542   specific language impairment: The role of contractibility and grammatical form

543   class,' *Journal of Speech, Language, and Hearing Research* 40/3: 480-492.

544   https://doi.org/10.1044/jslhr.4003.480

545   Cole, K. N., Mills, P. E., & Dale, P. S. 1989. 'Examination of test-retest and split-half reliability

546   for measures derived from language samples of young handicapped children,' *Language,*

547   *Speech, and Hearing Services in Schools* 20/3: 259-268. https://doi.org/10.1044/0161-

548   1461.2003.259

549   German, D. J. 1987. 'Spontaneous language profiles of children with word-finding

550   problems.' *Language, Speech, and Hearing Services in Schools* 18/3: 217-230.

551   https://psycnet.apa.org/doi/10.1044/0161-1461.1803.217

552   Grieser, D. L., & Kuhl, P. K. 1988. 'Maternal speech to infants in a tonal language: Support for

553   universal prosodic features in motherese,' *Developmental Psychology* 24/1: 14.

554   https://psycnet.apa.org/doi/10.1037/0012-1649.24.1.14

555   Hollister, J., Van Horne, A. O., & Zebrowski, P. 2017. 'The relationship between grammatical

556   development and disfluencies in preschool children who stutter and those who

557   recover,' *American Journal of Speech-Language Pathology* 26/1: 44-56.

558   https://doi.org/10.1044%2F2016_AJSLP-15-0022

559   Klusek, J., Martin, G. E., & Losh, M. 2014. 'A comparison of pragmatic language in boys with

560   autism and fragile X syndrome,' *Journal of Speech, Language, and Hearing*

561   *Research* 57/5: 1692-1707. https://doi.org/10.1044%2F2014_JSLHR-L-13-0064

562   Leadholm, B. J., & Miller, J. F. 1994. *Language Sample Analysis: The Wisconsin Guide.*

563   Bulletin 92424. https://eric.ed.gov/?id=ED371528

564   Leonard, L. B. 2014. *Children with Specific Language Impairment*. MIT press.

565   https://muse.jhu.edu/book/46919

566 Leonard, L. B., Eyer, J. A., Bedore, L. M., & Grela, B. G. 1997. 'Three accounts of the

567   grammatical morpheme difficulties of English-speaking children with specific language

568   impairment,' *Journal of Speech, Language, and Hearing Research* 40/4: 741-753.

569   https://doi.org/10.1044/jslhr.4004.741

570 Loban, W. 1976. *Language Development: Kindergarten through Grade Twelve.* NCTE

571   Committee on Research Report No. 18. http://files.eric.ed.gov/fulltext/ED128818.pdf

572 Miller, J., & Chapman, R. (2004). *Systematic Analysis of Language Transcripts (SALT v8.0).*

573   Madison, WI: Language Analysis Laboratory, Waisman Center, University of Wisconsin-

574   Madison.

575 Muma, J. R. 1998. *Effective Speech-language Pathology: A Cognitive Socialization Approach.*

576   Psychology Press. https://doi.org/10.4324/9781315805979

577 Nicholas, L. E., & Brookshire, R. H. 1993. 'A system for quantifying the informativeness and

578   efficiency of the connected speech of adults with aphasia,' *Journal of Speech, Language,*

579   *and Hearing Research* 36/2: 338-350. https://doi.org/10.1044/jshr.3602.338

580 Ott, B. L. 2017. 'The age of Twitter: Donald J. Trump and the politics of debasement,' *Critical*

581   *Studies in Media Communication* 34/1: 59-68.

582   https://doi.org/10.1080/15295036.2016.1266686

583 Pavelko, S. L., Price, L. R., & Owens Jr, R. E. 2020. 'Revisiting reliability: Using sampling

584   utterances and grammatical analysis revised (SUGAR) to compare 25-and 50-utterance

585   language samples,' *Language, Speech, and Hearing Services in Schools* 1:1-17.

586   https://doi.org/10.1044/2020_LSHSS-19-00026

587 Prins, R., & Bastiaanse, R. 2004. 'Analyzing the spontaneous speech of aphasic
588   speakers,' *Aphasiology* 18/12: 1075-1091. https://doi.org/10.1080/02687030444000534

589 Rice, M. L., & Wexler, K. 1996. 'Toward tense as a clinical marker of specific language
590   impairment in English-speaking children,' *Journal of Speech, Language, and Hearing*
591   *Research* 39/6 ; 1239-1257. https://psycnet.apa.org/doi/10.1044/jshr.3906.1239

592 Saffran, E. M., Berndt, R. S., & Schwartz, M. F. 1989. 'The quantitative analysis of agrammatic
593   production: Procedure and data,' *Brain and Language* 37/3: 440-479.
594   https://doi.org/10.1016/0093-934x(89)90030-8

595 Simon-Cereijido, G., & Gutiérrez-Clellen, V. F. 2007. 'Spontaneous language markers of
596   Spanish language impairment,' *Applied Psycholinguistics* 28/2: 317-339.
597   https://doi.org/10.1017/S0142716407070166

598 Solorio, T., Sherman, M., Liu, Y., Bedore, L., Pena, E., & Iglesias, A. 2011. 'Analyzing
599   language samples of Spanish–English bilingual children for the automated prediction of
600   language dominance,' *Natural Language Engineering* 17/3: 367.
601   http://dx.doi.org/10.1017/S1351324910000252

602 Tomasello, M., & Stahl, D. 2004. 'Sampling children's spontaneous speech: How much is
603   enough?,' *Journal of Child Language* 31/1: 101-121.
604   https://psycnet.apa.org/doi/10.1017/S0305000903005944

605 Tomblin, J. B., Harrison, M., Ambrose, S. E., Walker, E. A., Oleson, J. J., & Moeller, M. P.
606   2015. 'Language outcomes in young children with mild to severe hearing loss,' *Ear and*
607   *Hearing* 36: 76S-91S. https://doi.org/10.1097/aud.0000000000000219

608 Wagenaar, E., Snow, C., & Prins, R. 1975. 'Spontaneous speech of aphasic patients: A

609   psycholinguistic analysis,' *Brain and language* 2: 281-303.

610   https://psycnet.apa.org/doi/10.1016/S0093-934X(75)80071-X

611 Wittke, K., Mastergeorge, A. M., Ozonoff, S., Rogers, S. J., & Naigles, L. R. 2017.

612   'Grammatical language impairment in autism spectrum disorder: Exploring language

613   phenotypes beyond standardized testing,' *Frontiers in Psychology* 8: 532.

614   https://doi.org/10.3389/fpsyg.2017.00532

615 Wright, H. H., Capilouto, G., Wagovich, S., Cranfill, T., & Davis, J. 2005. 'Development and

616   reliability of a quantitative measure of adults' narratives,' *Aphasiology* 19/3-5: 263-273.

617   https://aphasia.talkbank.org/discourse/lit/Wrightetal2005.pdf

618 Zimmerer, V. C., Hardy, C. J., Eastman, J., Dutta, S., Varnet, L., Bond, R. L., ... & Varley, R. A.

619   2020. 'Automated profiling of spontaneous speech in primary progressive aphasia and

620   behavioral-variant frontotemporal dementia: An approach based on usage-

621   frequency,' *Cortex* 133: 103-119. https://doi.org/10.1016/j.cortex.2020.08.027

622 **Table 1**

623 *List and Examples of Target Morphosyntactic Structures*

Morphosyntactic Structures	Examples
multiverb utterances	he came and I ate (two clauses); it can drive fast (aux + verb)
copula	donkey is hungry
plural	I like spoons
present progressive (-ing)	Mommy is sleeping
genitive	it's Sophie's bear
contracted aux ('aux)	it's chasing the cow
do-aux: uncontracted	I do like it / do you like it
be-aux: uncontracted	they are going
third person singular (3s)	daddy eats cake
past tense	she walked; he ran

624

625 **Table 2**

626 *Frequency of morphemes in adult language samples*

Morphemes	Number of Instances			
	50	100	150	200
multiverb 1	27 (8.43)	52.3 (14.75)	75.95 (18.12)	89.75 (19.99)
multiverb 2	25.15 (4.60)	51.45 (9.74)	76.7 (13.43)	90.55 (15.34)
copula 1	14.9 (3.24)	30.45 (6.32)	45.45 (8.48)	53.5 (8.57)
copula 2	16.2 (3.52)	29.25 (5.51)	43.45 (7.88)	49.75 (8.75)
plural 1	10.6 (5.85)	21 (13.23)	28.5 (15.52)	33.1 (18.59)
plural 2	9.85 (3.91)	18 (8.45)	25.25 (9.81)	32.5 (18.27)
-ing 1	7.3 (4.10)	13.25 (7.22)	18.7 (9.99)	21.25 (11.17)
-ing 2	5.3 (2.58)	11.3 (4.78)	17.9 (6.15)	21.3 (6.28)
genitive 1	0.6 (0.75)	0.85 (0.99)	1.15 (1.46)	1.3 (1.69)
genitive 2	0.65 (1.18)	1 (1.62)	1.25 (1.97)	1.5 (2.14)
'aux 1	4.5 (4.16)	8.25 (5.62)	11.55 (7.52)	13.55 (7.81)
'aux 2	3.9 (2.40)	7.9 (5.06)	12.3 (7.62)	15.3 (9.91)
do-aux 1	4.9 (2.95)	9.65 (4.09)	13.6 (4.62)	15.5 (4.94)
do-aux 2	3.9 (3.37)	8.3 (4.37)	12.05 (6.64)	14.25 (6.90)
be-aux 1	5.05 (4.16)	9.8 (7.51)	13.5 (8.98)	15.85 (9.91)
be-aux 2	4.3 (2.41)	9.1 (4.45)	13.8 (6.39)	16.45 (7.18)
3s 1	3.5 (3.03)	6 (6.85)	7.85 (7.43)	8.65 (7.24)
3s 2	1.8 (1.73)	4.4 (4.04)	5.4 (3.86)	6.75 (4.41)
past 1	18.3 (13.39)	39.85 (19.5)	55.95 (24.75)	64.2 (25.32)
past 2	22.15 (13.36)	44.6 (19.73)	66.1 (30.28)	76.35 (32.19)

627

628 **Table 3**

629 *Number of specific morpheme types used by adults in 200 utterances in descending order of*

630 *frequency*

multiverb	90
past	70

copula	52
plural	33
-ing	21
be-aux	16
do-aux	15
'aux	14
3s	8
genitive	1

631

632 **Table 4**

633 *Reliability of adult language samples – Correlation between two samples*

	Number of Utterances			
	50	100	150	200
multiverb	.58*	.81*	.88*	.86*
copula	.65 *	.70*	.7*	.81*
plural	.49	.8*	.79*	.69*
-ing	.34	.56*	.48	.63*
genitive	.13	.35	.41	.47
'aux	.09	.82*	.84*	.89*
do-aux	-.32	.1	.08	-.11
be-aux	.44	.61*	.5	.51
3s	-.42	-.36	-.4	-.54
past	.38	.66*	.69*	.64*

634   *p < .05

635   **Table 5**

636   *Categories of reliability*

Unreliable:	genitive, do-aux, 3s
Partially Reliable:	-ing, be-aux, past
Reliable:	multiverb, copula, plural, 'aux

637

638   **Table 6**

639   *Frequency of Morphemes in Child Language Samples*

Morphemes	Number of Instances			
	50	100	150	200
multiverb 1	7.9 (4.3)	17.4 (8.3)	26.7 (10.8)	37.2 (15.2)
multiverb 2	11.0 (5.8)	21.8 (10.6)	32.7 (15.2)	44.3 (19.7)
copula 1	6.7 (4.5)	12.5 (7.2)	17.8 (10.4)	24.0 (13.0)
copula 2	5.7 (3.7)	11.3 (6.5)	15.9 (8.3)	20.1 (9.4)
plural 1	3.1 (2.7)	6.4 (2.9)	9.1 (3.6)	12.3 (4.8)
plural 2	2.5 (2.1)	5.7 (2.9)	8.7 (4.3)	12.1 (5.0)
-ing 1	2.1 (2.1)	4.5 (4.3)	7.4 (5.5)	9.8 (6.2)
-ing 2	2.9 (2.8)	5.8 (5.4)	8.7 (6.8)	11.0 (7.4)
genitive 1	0.22 (.60)	0.52 (.79)	0.91 (1.2)	1.1 (1.3)
genitive 2	0.17 (.39)	0.43 (.89)	0.61 (.94)	0.74 (.96)
'aux 1	1.8 (1.6)	3.9 (3.7)	6.1 (5.2)	9.0 (8.3)
'aux 2	2.4 (3.1)	4.5 (5.0)	6.1 (6.0)	8.3 (7.0)
do-aux 1	0.8 (1.1)	1.9 (1.7)	2.9 (2.1)	4.4 (2.9)
do-aux 2	1.3 (1.4)	2.5 (1.9)	3.3 (2.1)	5.3 (3.2)
be-aux 1	0.4 (0.9)	1.1 (2.4)	1.7 (3.2)	1.8 (3.2)
be-aux 2	0.4 (0.8)	0.9 (1.9)	1.3 (2.3)	1.4 (2.4)
3s 1	0.8 (1.3)	1.2 (1.3)	1.6 (1.7)	2.4 (2.7)
3s 2	1.2 (1.3)	1.8 (1.9)	2.1 (2.0)	3.0 (2.5)
past 1	0.8 (1.3)	2.1 (2.9)	3.3 (3.9)	4.5 (5.6)
past 2	0.7 (1.1)	1.6 (2.1)	2.8 (3.4)	3.6 (3.8)

640   *SD in parenthesis

641 **Table 7**

642 *Reliability of child language samples – Correlation between two samples in children aged 2;6-*

643 *3;6*

	Instances			
	50	100	150	200
multiverb	.56*	.70*	.70*	.73*
copula	.32	.64*	.70*	.63*
plural	.22	.24	.35	.49*
-ing	.26	.39*	.52*	.47*
genitive	.21	.12	.01	-.05
'aux	.29	.57*	.76*	.78*
do-aux	.27	.21	.41	.52*
be-aux	.34	.74*	.69*	.66*
3s	.16	.22	.36*	.38*
past	.11	-0.01	.32	.32

644

645 **Table 8**

646 *Morpheme Frequency Comparison of Adults and Children*

Adults		Children	
Multiverb	89.75	Multiverb	37.2
Past	64.2	Copula	24.0
Copula	53.5	Plural	12.3
Plural	33.1	-ing	11.0
-ing	21.25	'aux	9.0
Be aux	15.85	Past	4.5
Do aux	15.5	Do-aux	4.4
'aux	13.55	3s	2.4
3s	8.65	Be-aux	1.8

Genitive	1.3	Genitive	1.1	

647   Frequency using Time 1 at 200 utterances for adults and children ordered greatest to least