ROMA WIECZOREK
RWTH Aachen University

USING MLU TO STUDY EARLY LANGUAGE DEVELOPMENT IN ENGLISH

The study examines the parameter of Mean Length of Utterance (MLU), measured both in morphemes (MLUm) and words (MLUw), in early language development in the case of two English children matched for age. The MLU scores of a normally developing child were compared to the MLU results of a language-impaired child in a longitudinal study. Moreover, the reliability of the MLU index measured in words was also tested in both children. The MLU analysis was based on the CHILDES database and CLAN programme, where the transcripts of spontaneous speech samples are used to calculate basic language parameters at different age-points. The findings of this study indicate that despite the expected delay, the language-impaired child followed a similar route of language development as the control child. However, significant differences between MLUw and MLUm confirmed that the parameters performed two different linguistic analyses.

Key words: Mean Length of Utterance (MLU), Specific Language Impairment (SLI), morpheme, language development

Theoretical background

MLU as a quantitative measure of early language development

The notion of Mean Length of Utterance in morphemes (MLUm) was originally proposed by Brown (1973) and is in general related to Language Sample Analysis (LSA) which functions as an alternative to standardized tests like the Picture Peabody Vocabulary Test (PPVT) (DeThorne et al., 2005) or Northwestern Syntax Screening Test (Plante et al., 1993), for example. On the basis of a spontaneous speech sample or narrative provided by a child the number of morphemes is counted and then divided by the number of utterances. Originally, this relatively simple formula is assumed to correlate with the development of morphosyntactic skills in children and therefore can be assigned to a specific stage in language

Address for correspondence: Roma Wieczorek, Universitätsklinikum Aachen, Neurologische Klinik – Neurolinguistik, Pauwelsstr. 30, 52074 Aachen, Germany. E-mail: rwiezorek@ukaachen.de

Table 1. Stages of linguistic development according to Brown (1973)

Stage	MLUm value
I	1.75
II	2.25
III	2.75
IV	3.5
V	4.0

development (Brown, 1973). Brown's stages (1973) with corresponding MLUm values are illustrated in Table 1.

As such, the index of MLU is currently treated as a universal diagnostic instrument that signals potential language impairment. Firstly, it is believed to determine a child's stage of language development, as a high correlation with age was found (Miller & Chapman, 1981, Parker & Brorson, 2005). Secondly, it might help to identify and diagnose language deficit and measure possible improvements in language production (Paul, 2000). Finally, according to Bol (2003), the MLU score can be applied to contrastive studies as well, where the linguistic abilities of a group of language-disordered children are compared with those of typically developing children. As a result, MLU functions as a matching criterion and its value is kept similar in both tested groups. However, the author (Bol, 2003) points out that qualitative differences in the linguistic performance of control children and children with Specific Language Disorders (SLI) are often visible in the production of inflectional morphology, function words and personal pronouns. Furthermore, DeThorne et al. (2005) and Miller (1991) emphasize the role of MLU measuring only expressive language abilities, whereas the receptive aspect of a child's language can be better defined in terms of standardized tests. Thus, it is often claimed that as such, the MLU index should not replace other diagnostic instruments, but rather help to confirm the identification of language impairment (see also: Eisenberg et al., 2001).

Another potential problem with MLU refers to the language unit the index should measure. Apart from morphemes, there were also attempts to calculate MLU in syllables, but these were soon abandoned since the researchers encountered many problematic cases, such as frequent diminutives or duplication of syllables that appeared in children's speech (Bol, 2003, Parker & Brorson, 2005). The way MLU was calculated in words became disputable as well because of doubts concerning the selection of proper utterances. Namely, the issue of whether one- or two-word propositions should be treated as utterances or whether they should be excluded from the sample data remained open to debate (Parker & Brorson, 2005). On the other hand, the use of morphemes might contribute to the rapid growth

of the MLU scores in morphologically rich languages (Parker & Brorson, 2005). In addition, Eisenberg (2001) points to different rules for counting morphemes, which might not be universal and adaptable to diverse linguistic systems.

MLU and chronological age

Mean Length of Utterance is stated to be highly correlated to a child's chronological age. In consequence, the MLU value should reflect a specific stage in the child's language development and therefore estimate the actual age of the child (Miller, 1991). In other words, the child's age should grow together with language development. Although this assumption was confirmed in several studies (Klee, 1992, DeThorne, 2005), it was not supported by others, where the levelling off of the MLU index (Bol, 1996) or even its decline was demonstrated (Klee, 1992). Interestingly, the increasing tendency of MLU also seems to be inconstant (Bol, 1996). According to these results, MLU can be stated as not being fully dependent on age and its validity in correlation with chronological age can cease at some point. Thus, Blake et al. (1993), for example, suggest MLU validity for clausal complexity only up to the value of 4.5, whereas Miller (1991) extends the legitimacy of the measurement even for children at the age of 13, although no MLU values for the particular age-groups are provided.

Due to this inconsistency of results provided in recent studies, in the following experiment a normally developing child and a language-impaired child are matched for chronological age in order to observe potential MLU differences and alternations (see: Klee, 1992) as well as any significant differences between the MLU index measured in morphemes and in words. The chronological age as a matching criterion was chosen for one more reason. Namely, contrary to a language-age matching parameter, it prevents a situation in which younger normally developing children cannot participate in an experiment due to their inability to complete all the tasks (Plante et al., 1993; Bol, 2003). The second most important problem according to Plante et al. (1993) is the so-called "extraneous age effect." The authors (Plante et al. 1993) explain that, although significant differences in MLU values between language-impaired children and their age-matched peers are inevitable, the choice of language-matching procedure might bring the risk of finding no statistically significant difference in MLU scores between the control group and SLI children. Thus, it might result in the erroneous conclusion that language growth in both groups is "equal" (Plante et al., 1993; Bates, 1999).

In sum, apart from the weaknesses discussed above, Mean Length of Utterance remains one of the major indices used in evaluating and diagnosing children's language deficit and language development in general. Several attempts to find a substituting variable have been made so far, and one of them involved the application of MLUw, perceived by speech pathologists as a faster, easier and more adaptable parameter. MLU counted in words has already been applied to longitudinal studies, but to my knowledge, only in the case of children without language impairment.

Therefore, the current study will be devoted to finding the answer to the following research question concerning MLU:

– Is there a difference to be found between MLU counted in morphemes and MLU counted in words in a normally developing child and in a child with a language disorder?

Calculations of MLUm and MLUw on a limited number of 100 utterances

MLU in morphemes / words

The section below presents analyses of the two parameters, namely MLUm and MLUw, both measured in the case of two participants (boys) matched for chronological age (CA). The personal and linguistic data were obtained from the CHILDES database available online (Child Language Data Exchange System, http://childes.psy.cmu.edu/).

The samples in the case of both children were transcribed into morpheme-status (%MOR line), therefore the calculations of MLUw had to be done by hand. Thus, certain criteria (although subjective) had to be adopted:

– All contractions (e.g., *it's , what's, haven't*) were counted as <u>one</u> word
– If the *xxx* line appeared in the *CHI line, it was counted as <u>zero</u> words
– Onomatopoeic words (e.g., *uhm, wow*) were counted each as <u>one</u> word
– All the repetitions of one word (e.g., *no, no, no* → 3 words) were counted separately.

The data below encompass 18 files altogether (9 for each child) and reflect the linguistic performance of the children in the 3;5 – 4;5 yrs time frame. In the case of the SLI child, the first file contained the transcript of only 84 utterances. However, despite the insufficient number of utterances, it was adopted for the task in order to observe and comment upon any possible difference with the adequate file taken from the other boy's database. The statistical analysis was done by means of a series of nonparametric tests performed using SPSS software (Moore & McCabe, 2003).

Subjects

a) Harry – SLI child

The child was diagnosed as an SLI speaker and received language therapy during the recording session. By the end of the data collection period, Harry was enrolled in a mainstream primary school. The study encompassed a period of 15 months, with Harry being 3;5 yrs at the beginning and 4;8 yrs at the end of the session. However, the files used to analyze the values of MLU in words and morphemes comprised only the age-period from 3;5 yrs to 4;5 yrs. The child's background showed no history of hearing problems, a good degree of intelligibility, non-verbal abilities within the normal range, no obvious autistic tendencies, but poor language abilities

Table 2. Conti-Ramsden 3 – Harry's data (adapted from CHILDES Manual Database); MLUw – Mean Length of Utterance in words; I.Q. – Intelligence Quotient

Child	Age	MLUw	I.Q.
Harry	3;5.0	1.78	120

including poor receptive abilities. The longitudinal data presented were derived from the Conti-Ramsden-3 folder found in the CHILDES database.

b) Stuart – ND child

This directory contains transcripts from a study of eight upper-working-class children acquiring English in Belfast, Northern Ireland between 1995 and 1998. The data on Stuart's linguistic performance are based on 9 files taken from the Belfast folder representing the collection of British English spontaneous speech samples of normally developing children. At the beginning of this longitudinal study Stuart was 3;5 yrs old and the last sample was recorded when the child was 4;5 yrs old.

Table 3. Belfast directory – Stuart's data (adapted from CHILDES Manual Database)

Child	Birth date	Age
Stuart	21-APR-1992	3;5.12

Results

General analysis

a) Harry – SLI child

There are two interesting patterns that are well visible in Harry's MLUm and MLUw data. As Figure 1 shows, the values for the two parameters grow from the approximate age of 3;7 yrs till the age point 3;11. Then, the "morpheme" line seems to be quite stable and oscillates around the value of 4 (apart from a sudden drop at the age point of 4;11 yrs). In the case of the "word" analysis, though, the values fluctuate. Still, the overall patterns of MLUw and MLUm are similar as the chronological age rises. The only difference is noted in the maximum values the indices represent at the age of 4;5 yrs, namely the MLUm score is higher (around 4) and the MLUw rate does not exceed the value of 3.0.

b) Stuart – ND child

When it comes to Stuart's data, it might be noticed that the "morpheme" line has its constant fluctuations, whereas the "word" line keeps increasing steadily from the age of 3;7 yrs until the age of 4;4 yrs. Contrary to the "morpheme" line, which maintains the value of 5 at the age point of 4;4, the "word" line plummets from the value of 5 to approximately the value of 3.5.

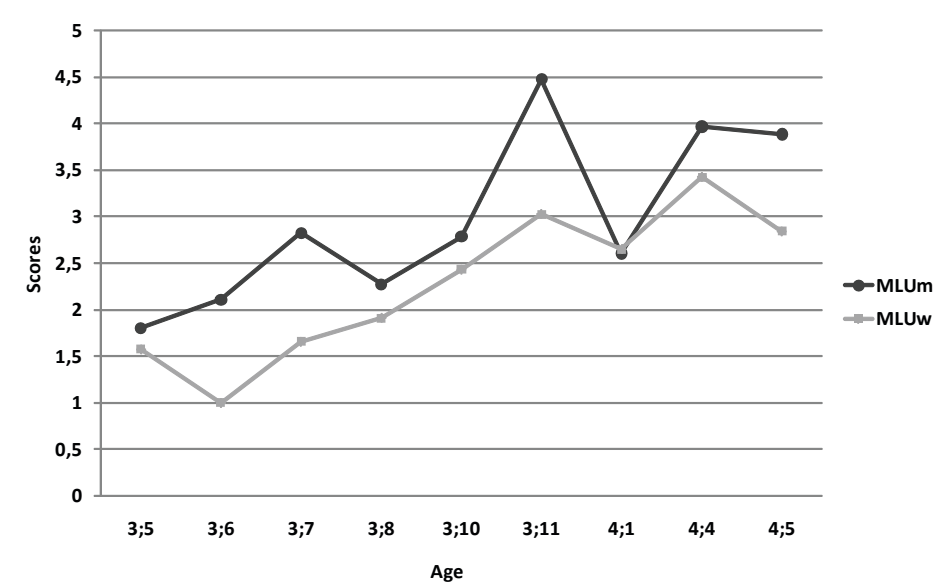Figure 1. Comparison of MLUm and MLUw for Harry (SLI child)

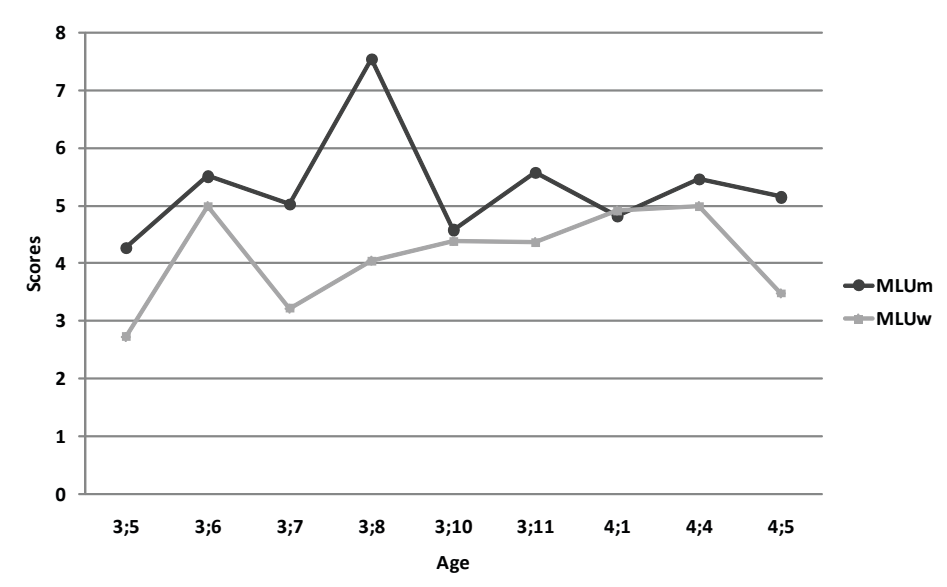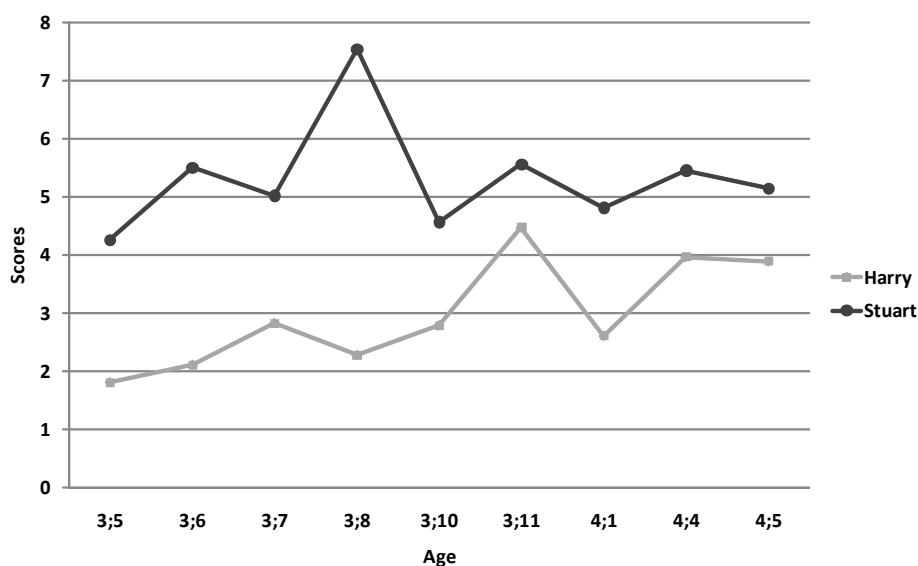Figure 2. Comparison of MLUm and MLUw for Stuart (ND child)

Figure 3. Comparison of MLUm scores

c) Harry vs. Stuart

Figures 3 and 4 illustrate the children's performance with respect to the two parameters tested. As regards the MLUm condition, in the case of Stuart, the control child, the greatest increase occurs around the age point of 3;8 yrs, whereas for Harry, the SLI child, this point is delayed till 3;11 yrs. Secondly, a sudden drop of MLUw scores for Harry at the age of 3;6 can be observed and at the same time Stuart reaches the peak of his MLUw rate. However, it can be noticed that, although the language-impaired child achieves worse results than the control child at the same age, the overall pattern of MLUm and MLUw performance for both children is similar. Therefore, it might be concluded that probably due to therapy, Harry's linguistic behavior resembles Stuart's, but the language development of the SLI child is nevertheless affected as Harry never reaches Stuart's scores.

Statistical analysis

In order to find any significant quantitative differences between the MLUw and MLUm rates observed in the study, a Mann-Whitney nonparametric test was carried out. First of all, the test revealed that the MLUm values were significantly different between the normally developing child and his SLI peer: $Z = -3,488$; p (1-tailed) < 0.000. Moreover, the results from MLU counted in words were also compared and showed a significant difference as well. Namely, Harry's outcome was again significantly lower ($Z = -3,225$; p (1-tailed) < 0.0005.

Figure 4: Comparison of MLUw scores
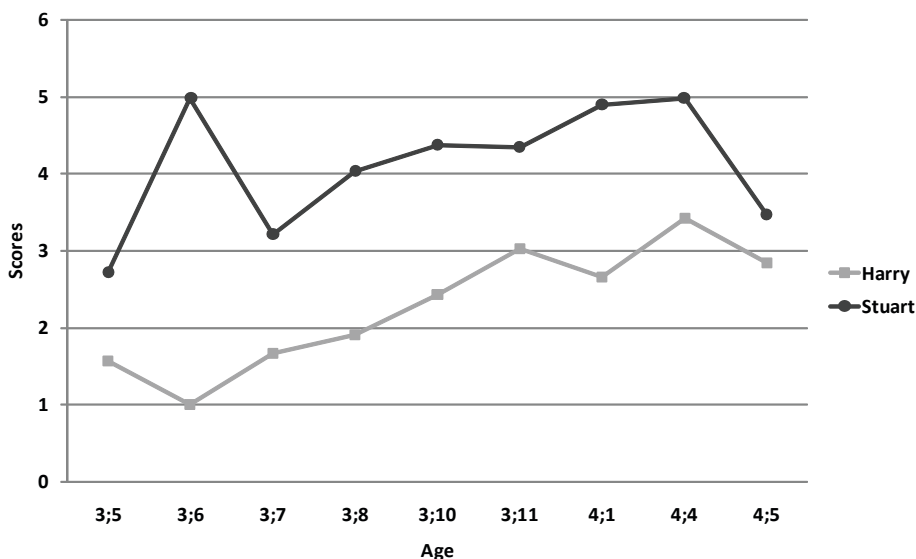
Apart from the analysis *between* the participants, we also took an interest in the comparison (MLUm/MLUw) *within* the subjects. This calculation was conducted using the Wilcoxon Signed Rank Test. It turned out that the values recorded for MLUm and MLUw differed significantly in the case of both children, i.e. the MLUw score was lower than the MLUm score: $Z = -2,521$; p (1-tailed) < 0.006 and $Z = -2,666$; p (1-tailed) < 0.004 for Harry and Stuart, respectively.

Finally, a correlation analysis between MLUm and MLUw was performed and showed significant values for each participant ($r = 0.84$ for Harry and $r = 0.61$ for Stuart at the 0.01 and 0.05 level (1-tailed), respectively).

Discussion

MLU in words vs. MLU in morphemes

The results presented above suggest a high correlation between MLU calculated in words and MLU measured in morphemes for each of the subjects. It is worth noting that similar results were obtained in the study conducted by Parker & Brorson (2005) on normally developing children, when the correlation value of 0.99 and $p < 0.001$ was illustrated, suggesting that MLUw can be considered as an alternative measure to MLUm (see also: Bol, 2003). Furthermore, similar patterns of the children's behaviour on both MLUm and MLUw were noted, although the MLU scores obtained

by Harry reflect only the attempt to "keep up with" Stuart's development. Hence, Johnston & Kamhi (1984:65) argue that "[a]lthough language-impaired children learn to talk slowly and late, they learn phrase structure rules, morphological forms... in sequences which can be predicted from research with normal children." In other words, delayed as the acquisition of language might be in SLI children, it follows the same route of development as in control children (see also: Johnston, 2001).

However, the significant difference found between MLUw and MLUm scores might also indicate that the two indices reflect two different types of linguistic development. Namely, MLUw provides information about the lexicon, whereas MLUm illustrates morphosyntactic development. Therefore, it is also claimed that the high correlation between these parameters is not enough to substitute one for the other, as Peters (1999:1) emphasizes that "on the one hand there is the stringing together of content words (MLUw) and on the other the increasing inclusion of grammatical markers (MLUm)." Thus, the author concludes that these two phenomena have to be assessed separately (ibid.). Indeed, it was observed in the current study that already in this early language development the quantitative difference between words and morphemes could have been marked. Therefore, a more qualitative attitude towards the research is advocated, that is, apart from knowing *how* to use statistics we should also pay attention to *what* is actually being measured (Bol, 2003; Leonard & Finnerman; 2003, Peters, 1999). In sum, the relatively high correlation between the two indices may signal that both MLUm and MLUw variables reflect well the particular stages of early language development. However, it cannot be ignored that the two indices are responsible for measuring two different linguistic aspects. Thus, in order to check whether the difference between parameters remains as children grow, it would be interesting to study the relation between MLUw and MLUm also in older children.

References

Bates, E. (1999). Letter to the Info-Childes. *Child Language Bulletin*, 19 (1).

Blake, J., Quartaro, G., & Onorati, S. (1993). Evaluating quantitative measures of grammatical complexity in spontaneous speech samples. *Journal of Child Language* 20, 139-152

Bol, G.W. (1996). Optional subjects in Dutch child language. In Ch. Koster & F. Wijnen (Eds.), *Proceedings of the Groningen Assembly on Language Acquisition* (pp. 125-133). Groningen: Centre for Language and Cognition.

Bol, G.W. (2003). MLU-Matching and the Production of Morphosyntax in Dutch Children with Specific Language Impairment (SLI). In Y. Levy & J. Schaeffer (Eds.), *Language Competence Across Populations: Toward a Definition of Specific Language Impairment* (pp. 259-272). Mahwah, NJ: Lawrence Erlbaum.

Brown, R. (1973). *A first language: The early stages.* Cambridge, MA: Harvard University Press.

DeThorne, L.S., Johnson, B.W., & Loeb, J.W. (2005). A closer look at MLU: What does it really measure? *Clinical Linguistics and Phonetics,* 19 (8), 635-648.

Eisenberg, S.L., Fersko, T.M., & Lundgren, C. (2001). The Use of MLU for identifying language impairment for preschool children: A review. *American Journal of Speech-Language Pathology*, 10, 323-342.

Johnston, J.R. (2001). An alternate MLU calculation: Magnitude and variability of effects. *Journal of Speech, Language and Hearing Research*, 44, 156-164.

Johnston, J.R. & Kamhi, A. 1984. Syntactic and semantic aspects of the utterances of language impaired children: The same can be less. *Merrill-Palmer Quarterly*, 30, 65-85.

Klee, Th. (1992). Measuring Children's Conversational Language. In S.F. Warren & J. Reichle (Eds), *Causes and effects in communication and language intervention* (pp. 315-330). Baltimore: Brookes.

Leonard, L.B. & Finnerman, D. (2003). Grammatical morpheme effects on MLU: "The same can be less" revisited. *Journal of Speech, Language and Hearing Research,* 46, 878-888.

Miller, J. (1991). Quantifying Productive Language Disorders. In J. Miller (Ed.) *Research on Child Language Disorders: A Decade of Progress* (pp. 211-220). Austin, TX: Pro Ed.

Miller, J.F. & Chapman, R.S. (1981). The relation between age and mean length of utterance in morphemes. *Journal of Speech and Hearing Research*, 24, 154-161.

Moore, D.S. & McCabe, G.P. (2003). *Introduction to the Practice of Statistics.* New York: Freeman.

Parker, M.D. & Brorson, K. (2005). A comparative study between mean length of utterance in morphemes (MLUm) and mean length of utterance in words (MLUw). *First Language,* 25 (3), 365-376.

Paul, R. (2000). *Language Disorders from infancy through adolescents* (2nd ed.). Saint Louis, MO: Mosby-Year Book.

Peters, A. (1999). From Info-Childes. *IASCL, Child Language Bulletin*, 19 (1).

Plante, E., Swisher, L., Kiernan, B., & Restrepo, M.A. (1993). Language Matches: Illuminating or Confounding? *Journal of Speech and Hearing Research*, 36, 772-776.

Retherford, K.S. (1993/2000). *Guide to analysis of language transcripts* (2nd ed.). Eau Claire, Wl: Thinking Publications.

SPSS Inc. (2001). SPSS (Version 11.0). [Computer software]

Appendix 1. Rules for counting morphemes (Retherford, 2000)

1. Fillers are not assigned morphemes (*um, well, oh)*
2. Compound words and closely related ones are assigned 1 morpheme (*all gone, bye-bye*)
3. Indefinite and reflexive compound pronouns are assigned 1 morpheme (*herself, anything*)
4. Proper nouns and ritualized reduplications are assigned 1 morpheme (*Mr. Smith, night-night*)
5. Diminutive forms of words receive 1 morpheme (*funny, doggie*)
6. Auxiliary verbs are assigned 1 morpheme
7. Catenative forms are assigned only 1 morpheme (*gonna, wanna, hafta*)
8. Inflectional affixes (i.e. plural *−s* , singular and plural possessive-*s*, present third person singular -*s*, regular past tense -*ed* and *−en*, present participle *−ing*, comparative *−er* , and superlative *−est*) are assigned morphemes
9. Incorrect uses of inflection are not counted as separate morphemes
10. Irregular past tense and past participle forms are assigned only 1 morpheme
11. Negative contractions are assigned two morphemes if there was evidence within the transcript that the child used each part of the contraction separately. If the child does not use each part of the contraction separately, 1 morpheme is assigned
12. All nonnegative contractions are assigned 2 morphemes
13. Common derivational affixes are assigned their own morphemes

Appendix 2. CLAN commands used in the essay

```
mlu +t*CHI sample.cha
mlu +z100u-200u +t*CHI sample.cha
freq +t*CHI sample.cha
freq +z100w-200w +t*CHI sample.cha
kwal +z100w +t*CHI sample.cha
kwal +z100w +t*CHI sample.cha >> output100w
ren output100w stu10.cha
vocd stu10.cha
```