**On the L1-Acquaition of variation in adjective amplification**

Martin Schweinberger

The University of Queensland, Australia

School of Languages and Cultures

Gordon Greenwood Building

St. Lucia Queensland 4072

Australia.

m.schweinberger@uq.edu.au

**Abstract**

This corpus-based study analyzes the L1-acquisition of adjective amplifiers in American English based on data from the *Child Language Data Exchange System* (CHILDES). To investigate how monolingual children acquire lexical variants in a variable context (adjective amplifiers such as *very*, *really*, and *so*) and what factors affect the acquisition of, the study adopts a multifactorial deviation analysis based on random forest classifications (MuPDARF) to analyze whether the output of children aligns best with, and is thus most similar to, (i) the input of caregivers (child-directed speech of mothers and fathers; CDS), (ii) siblings, or (iii) the communal grammar (non-child directed speech of men and women in the same age of the mother and fathers). The statistical analysis shows that the use of adjective amplifiers is exhibits a remarkable degree of systematicity and that children as young as 4-years old are sensitive to language-external factors (the situational setting of conversations). In addition, the output of children aligns most closely with the input of siblings, thus showing that even at very early stages of acquisition, children mirror peers rather than the input of mothers or the wider communal grammar. The results are interpreted to support the hypothesis that systematic variation is acquired alongside linguistic constructions themselves. It is argued that the early acquisition of variation supports basic tenants of usage-based theories of language acquisition according to which linguistic constructions are learned in a socially grounded manner and discourse structures which are frequently used in unplanned discourse are acquired particularly early during L1 acquisition.

1 **Introduction**

L1-acquisition research addresses the question of how children acquire the language that is in use around them (Clark 2016: 18). Research on this issue has as produced a substantive amount of studies which provide a very detailed picture of the mechanism that underly the acquisition process as well as the developmental stages that children go through and the linguistic abilities, errors, and features that characterize these stages (see Clark 2016 for an overview). But, given that children encounter a myriad of different options for how to express themselves in the input they receive, questions as to the impact and importance of factors that differentiate the input itself but also set distinct types of input apart remain unresolved (Nardy, Chervot, & Barbu, 2013: 256).

While there is a growing body of research in variationist sociolinguistics on pre-adolescents and the acquisition of variation during very early stages of L1-acquisition (see especially Smith & Durham 2019), issues relating to the both the interplay of factors at work in the acquisition of variation and indeed the acquisition of variants within a variable context remain under-explored (Nardy, Chervot & Barbu, 2013: 256). Early theoretical models dealing with the acquisition of variation have assumed that children remain monolectal until approximately age ten and only then acquire systematic extra-linguistic constraints based on situational setting (Labov 1964; cf. Nardy, Chervot & Barbu, 2013: 258-260). In contrast, recent studies dealing with the acquisition of variation triggered by extra-linguistic factors (for instance, Smith, Durham & Fortune 2007; Smith, Durham & Richards 2013; Smith & Durham 2019) contradict this model and show that variation is acquired very early on – probably alongside the acquisition of linguistic constructions themselves. The present study aims to address this issue with regard to adjective amplification.

XXX include examples here XXX

Words are not only fundamental to the acquisition of any language but children also pick up words like magnets pick up pins, as famously stated by Aitchison (2012: 209). When acquiring lexical elements (words), children do not only have to acquire their semantic meaning and syntactic properties but also their “appropriate”, sociolinguistic use and their covert as well as overt prestige (Trudgill 1972). Yet, and despite their importance, studies which have focused on the acquisition of variation have mostly addressed the acquisition of phonological variants but, so far, only very few studies have addressed variable use of lexical variants in a variable context (see Smith & Durham 2019 as a notable exception). Research on the acquisition of phonological variation (for instance, Roberts, 1997; Roberts & Labov, 1995; Smith, Durham & Fortune, 2007; Smith, Durham & Richards, 2013; for an overview cf. Nardy, Chervot & Barbu, 2013: 260) have investigated phenomena such as liasion in French (Barbu, Nardy, Chervot & Juhel, 2013), vocalic differences (Macaulay 1977), (-t,d) deletion (Roberts, 1997), (ing) realization (Roberts, 1996), the production of word-final (r) (Romaine 1978) or short (a) (Roberts & Labov, 1995). These studies have shown that even children aged three and four are able to acquire phonological variants along with the respective constraints on stable variation. To elaborate, Smith, Durham & Fortune (2007) showed that children from 3;2 onwards use more non-standard variants in informal settings such as daily activities or game play compared to interactions of an educational or disciplinary character and Roberts & Labov (1995) found that children from Philadelphia aged three and four accurately acquired phonological variants, short a, along with the respective constraints and concluded that even very young speakers take part in ongoing language change.

Studies that have explicitly focused on the acquisition of lexical variants in a variable context have found that the output of children aligns very closely with the input of caregivers (Smith & Durham 2019: 91) – at least with respect to frequency of use. Both the CDS and the children’s output differ drastically from usage patterns in the wider community. However, both children and CDS approximates the community norms as children grow older and their output thus increasingly conforms with the communal grammar (Smith & Durham 2019: 93). With respect to gender, the analyses presented in Smith & Durham (2019) did neither find significant differences with respect to the children’s nor with respect to the caregivers’ gender. In addition, only older children were apparently aware of situational constraints on language sue as significant differences among more and less formal settings could only be confirmed for older but you younger children (). It is important to note, though, that caregivers of young children were more conservative in their CDS and exhibited lower frequencies of nonstandard, local lexical variants compared with standard forms (*aye* versus *yes*) (see Smith & Durham 2019: 93).

Despite this growing amount of knowledge on the L1-acquisition of lexical variants, there exist at least two areas that require further exploration: firstly, while recent studies on phonological variation (e.g. Smith, Durham & Fortune, 2007; Smith, Durham & Richards, 2013; Roberts, 1997; Roberts & Labov, 1995) have found that children acquire patterns of variation and the respective constraints already between the ages of three and four., Smith and Durham (2019: 93) showed that only older children were aware of situational constraints when it comes to lexical variants.

The finding that stylistic variation is acquired alongside the acquisition of linguistic constructions is interesting as it substantiates hypotheses derived from usage-based approaches on L1-acquisition (cf. Behrens, 2009; Ibbotson, 2013; Tomasello, 2003; Tomasello 2012 for an overview) which assume that children acquire extra-linguistic factors and constraints are learned together with linguistic constructions (cf. Hilpert 2014: 159; cf. also Tomasello, 2012 and Hilpert 2014: 155-178 for an overview). Secondly, previous research has claimed that DM like is acquired rather late, i.e. after an age of ten (Miller & Weinert, 1995). However, if DM like is acquired that late, then this would not only be at odds with predictions that can be derived from a model of discourse production proposed by Ochs (1979) but also, as pointed out above, with basic theorems of usage-based approaches on L1-acquisition which assume that linguistic constructions with occur frequently in the input of the child, i.e. in the child directed speech produced by the mother, are acquired very early on. Thus, usage-based approaches stress the importance of the input – typically produced by the mother – for the time of acquisition. Ochs’ (1979) model is relevant here as it argues for what is today known as entrenchment in usage-based models of language production and acquisition (cf. e.g. Schmid 2007; Lieven and Tomasello 2008). According to Ochs (1979:53) speakers rely on morphosyntactic and discourse skills acquired between the age of three and four when engaging in unplanned discourse (for a more recent elaboration of this point from the perspective of usage-based language acquisition cf. Lieven & Tomasello 2008). However, Miller & Weinert (1995), who were among the first to investigate the use of DM like by pre-adolescents, argue that this is incorrect with respect to DM like: while DM like is highly frequent in unplanned discourse, Miller & Weinert (1995:366) conclude that it is apparently acquired after the age of ten based on data from The Hcrc Map Task Corpus (Anderson, Clark & Mullin, 1991). Unfortunately, Miller & Weinert (1995) as well as later studies on the use of DM like (e.g. Levey, 2006) in pre-adolescent speech use data of speakers unfit to investigate the use of DM like during early stages of L1-acquisition and to investigate the acquisition of situational constraints.

The present study aims to close this gap in our knowledge about the acquisition of variation based on the DM like by investigating its use in the speech of children aged between three and a half and twelve years of age based on the Home-School Study of Language and Literacy Development (HSLLD) component of the Child Language Data Exchange System (CHILDES) (MacWhinney, 2000). In this respect, the present study addresses the following research questions:

1. At which age is the DM like actually acquired?
2. Which factors correlate with the use of the DM like?
3. Do children acquire contextual, i.e. stylistic, constraints alongside the acquisition of DM like?
4. Does the use of the DM like of the mother (i.e. the assumed primary caregiver) correlate with the age at which a child acquires DM like, i.e. do children whose mother's use DM like more acquire DM like earlier?

The following section describes the data and its processing for the present analysis. The results of the present study are presented in section 4 and section 5 discusses these findings and offers an outlook on future research.

3 Data and Methodology

The following section describes the corpus data used in this study, details the coding of the variables, i.e. the operationalization of the variables, and provides information on the statistical analysis, i.e. the type of regression model and the process of model-fitting used in this study.

3.1 Corpus selection

Despite a growing amount of sophisticated corpus-based analyses of the acquisition of linguistic variation (e.g. Smith, Durham & Fortune, 2007; Smith, Durham & Richards, 2013), previous studies of child language variation have often relied on small data sets or even diary entries due to a lack of sufficiently large data sets (Roberts, 2008). To provide a fine-grained analysis not only of the linguistic output of children during L1 acquisition but also to determine the role of the input of the mothers, the present study utilizes the Child Language Data Exchange System (CHILDES) first conceived in 1981 (MacWhinney, 2000:8). The CHILDES contains various sub-corpora of which The Home-School Study of Language and Literacy Development corpus, short HSLLD, is particularly suited for the present investigation[[1]](#endnote-1) as (i) it represents speakers between the ages of three and a half and twelve, (ii) it contains a sufficiently large number of speakers, (iii) it is sufficiently large in terms of quantity of words, (iv) it represents speech produced in controlled environments and settings, (v), and it represents data collected from the same children over a significant time span and thus allows to track the linguistic development of children.

The HSLLD data have been collected between 1988 and 1996 and contain 278,895 words (counting only children’s contributions, cf. Table 1 for a more exhaustive overview). Children were recorded at home in presence of their family and the investigator at approximately ages 3, 4, and 5, as well as in 2nd and 4th grade[[2]](#endnote-2). Each visit lasted between one and three hours and consisted of a number of tasks:

1. book reading (The Very Hungry Caterpillar)
2. elicited report (children were asked to share some past experience)
3. letter writing (to the author of The Very Hungry Caterpillar)
4. experimental (magnet) task
5. mealtime conversations (in absence of the investigator)
6. toy play (including blocks, toy cars, and puzzles brought by the investigator)
7. reading (reading an unspecified text)
8. mother definitions (removed from the data as it was only conducted during the last home visit and information about what this situation type consisted of has not been made available in the respective manual (cf. CHILDES:64))

Although it is very fortunate that the HSLLD data represents distinct situation types, these situation types differ substantially with respect to the quantity of words uttered in these situations by the children (Figure 1) and they were therefore collapsed into the two categories formal and informal in the final data set (Figure 2) (please note, that the figures below show the data of only those 46 children that were part of the final data set).

|  |  |
| --- | --- |
| C:\03-MyProjects\06LIKEAcquisition\Article\images\stypwcpcnt.png  Figure 1 Absolute number and percentages of words by situation type in the HSLLD corpus | C:\03-MyProjects\06LIKEAcquisition\Article\images\stypwcpcnt2.png  Figure 2: Absolute number and percentages of words by situation type in the recoded data set |

Initially, data was collected from 85 children but this number had shrunk to 74 at age 5 and dwindled to 68 by the time children were in 2nd grade and only 46 children partook in all visits and situation types (as will be stated below, only these 46 children remained in the final data set). Nearly all children were raised in low-income families[[3]](#endnote-3) and grew up in or around Boston, Massachusetts. The corpus is evenly balanced among the sexes (36 girls and 32 boys) and includes 16 African American (21.5%), 6 Hispanic (8.1%), and 5 biracial (7%) families who participated at least till age 5 (percentages are based on the number of participating families at age 5) (cf. CHILDES: 62-63). Roughly, one third of mothers are high school dropouts, one third received only a high school diploma, and one third had attended some post-high school education. It needs be mentioned that, unfortunately, many of the socio-linguistic variables like ethnicity, education of the mother, and family income are (for reasons of confidentiality) not available for individual families, but were used to establish the general make-up of the corpus (but not published).

3.2 Data extraction and processing

Extracting all instances of the orthographic sequence <like> yielded 9,598 hits. As these 9,598 hits contain both instances of DM like as well as instances of non-discourse marker like, each data point was manually coded as being either an instances of DM like or of non-discourse marker like. Next, the visit, speaker, and situation type were extracted for each instance of DM like and added to the table. Next, the table in which each row represented one instance of DM like was reorganized so that each row would now represent the number of DM like per child per situation type and per visit. Thus, each row represented the absolute frequency of DM like per child and situation type and visit. Furthermore, the relative frequency of DM like by the mothers’ irrespective of visit or situation type were calculated and added to the data table. In a next step, biographical parameters of each child in any of the given visits and situation types, e.g. its age and gender, were extracted and their respective word counts were calculated. The children's biographical information was then combined with the tabulated instances of DM like per child and situation type and visit. Based on this resulting table, it was determined whether the child was falling into the category of DM like user or not in a given home visit and in a given situation type. Finally, only those children that took part in all five home visits and were recorded during all situation types remained in the data, i.e. the data this study is based on represents 46 children who were visited five times while all other files from children and interlocutors that did not fulfill these criteria were removed from the analysis. Before turning to the operationalization of the variables, the following subsection will elaborate on the coding of individual instances and functions of the orthographic sequence <like>.

3.3 Coding and classification

This section will briefly discuss how the individual instances of orthographic sequence <like> were coded as to illuminate which of the various functions of the orthographic sequence <like> were considered instances of the DM like.

3.3.2 Discourse marker LIKE

The orthographic sequence <like> was coded as an instance of DM like if that instance was syntactically optional, semantically bleached, and its removal did not substantially affect the acceptability of the remaining sentence (cf. (1)). Instances of <like> which represented the use of <like> as a transitive verb meaning to fancy, the use of <like> as a comparative preposition (cognate with modern German gleichen) which takes a nominal or a sentential complement were coded as non-discourse marker uses of <like> and accordingly removed from the analysis.

While the above stated classification was sufficient to classify the vast majority of instances, there were borderline cases which require additional attention.

One such case occurred when <like> preceded numerals (10a) and occasionally demonstratives (10b). In such cases, <like> is a borderline case between discourse marker and adverbial (Andersen, 2001:260) and may be substituted by traditional approximants like about, roughly, or approximately. Such instances were considered representations of DM like in the current study as their removal would render the remaining structure ungrammatical.

1. a. \*CHI: at like twelve they had the little slash and I +... (HV7/MT/conmt7)

b. \*CHI: oh it's only like that (.) big . (HV7/MD/jammt7)

In a very similar case DM like functions as a focalizer, namely, when it precedes a numeral which introduces or is part of an exaggeration as in (11).

1. \*CHI: was it [= the wedding cake] like one mile high ? (HV5/MT/jebmt5)

These instances were classified be being instances of DM like because - even though they impact the truth conditions of the underlying sentence (cf. Siegel, 2002) - they co-occur with traditional adverbs of approximation and can be left out without rendering the remaining utterance ungrammatical.

The following section discusses the operationalization of the variables investigated in the present study.

3.4 Operationalization of variables

The variables represent information contained in the data itself, e.g. the use of DM like by a child in a given situation type during a certain visit or the overall use of DM like by a child's mother, or they represent biographical information about the children, e.g. the age and gender of a child. The source for the biographical information about the children is meta-information provided in the headers of the HSLLD files. The following subsection describes the variables and their operationalization in greater detail to enable replication.

3.4.1 Use of discourse marker like (likeyn dependent variable)

The dependent variable in this study is the use of DM like in transcribed speech of a child in a certain situation type during a given home visit. Each file representing a situation type during one of the home visits was coded as 0 (no occurrences of DM like in speech of the child) or as 1 (DM like was present in the speech of the child). The resulting factor has accordingly two levels (0, 1) and represents a nominal variable.

3.4.2 Age (ageedit, independent variable)

The age classification of the HSLLD is quite fine-grained and provides the year, number of months and days the child was old when the recordings took place. For the current study these ages were converted to decimal numbers with days being removed, i.e. a child that was 3 years, 6 months and 17 days old (3;06:17) would be assigned an age of 3.5, while a child that was 4 years, 9 months and 8 days old (4;09:08) would be assigned an age of 4.75. Furthermore, all 6 conversations in which children were older than 10 years of age were removed from the data set because their inclusion would have over-proportionately affected the results of the regression analysis and thereby rendered these results less reliable. The age of children was accordingly converted into a continuous numeric variable that ranges between 3.5 (3 years and 6 months) and 10.[[4]](#endnote-4)

3.4.3 Gender (sex, independent variable)

Although gender and sex are two distinct concepts, the classification of gender is based on the reported information provided in the headers of the HSLLD files. In the present study, gender is a nominal variable representing a factor with two levels (female, male).

3.4.4 Situation type (styp, independent variable)

Originally, the data contained seven situation types: toy play (tp), book reading (br), elicited report (er), experimental task (et), reading (re), letter writing (rlw), mother definitions (md), and meal time (mt). However, conversations falling into the mother definitions (md) category were removed from the data as it was only conducted during the last home visit and information about what this situation type consisted of has not been made available in the respective manual. In addition, as these situation types differ substantially with respect to the quantity of words uttered in these situations by the children (Figure 1) and they were therefore collapsed into the two categories formal and informal in the final data set (Figure 2). The coding of a situation type depended on the information provided in the headers of the HSLLD files. The resulting factor has two levels (formal, informal) and is a categorical variable.

3.4.5 Mothers’ use of the discourse marker like (motlihvstptw, independent variable)

The mothers’ use of DM like represents the relative frequency of DM like per 1,000 words in a given situation type during one of the home visits of the mother. Accordingly, the resulting variable is numeric.

3.5 The final data set

The following section summarizes and displays the properties of the final data set, i.e. the data that the analysis of DM like has been based on. An overview of DM like use by children, mothers and other interlocutors is provided in Table 1.

|  |  |  |  |
| --- | --- | --- | --- |
|  | LIKE | Words | Rel. Frequency |
|  | (N) | (N) | (ptw) |
| Children | 232 | 278,895 | 0.832 |
| Mothers | 296 | 753,939 | 0.393 |
| Others | 266 | 234,024 | 1.137 |
| Total | 794 | 1,266,858 | 0.630 |

Table 1: Summary of the instances of DM like by speaker cohort

Table 2 displays the data with respect to the number of speakers, words, DM like-users, and genders within each sub-cohort.

Table 2: Summary of the final data set with respect to age and gender

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Age | Sex | Speakers | DM like | Words | Mean DM like | Percent |
|  |  | (N) | (N) | (N) | (ptw) | (DM like user) |
| 3 | female | 12 | 0 | 15,645 | 0.000 | 0.00 |
| 3 | male | 12 | 6 | 15,175 | 0.517 | 25.00 |
| 4 | female | 24 | 10 | 36,480 | 0.151 | 25.00 |
| 4 | male | 21 | 10 | 33,865 | 0.238 | 25.00 |
| 5 | female | 23 | 13 | 40,497 | 0.457 | 28.12 |
| 5 | male | 23 | 11 | 35,074 | 0.414 | 25.81 |
| 6 | female | 4 | 0 | 2,354 | 0.000 | 0.00 |
| 6 | male | 3 | 1 | 2,344 | 0.236 | 33.33 |
| 7 | female | 23 | 17 | 19,978 | 0.616 | 30.30 |
| 7 | male | 21 | 25 | 18,747 | 0.48 | 23.08 |
| 8 | female | 1 | 0 | 1,800 | 0.000 | 0.00 |
| 8 | male | 0 | 0 | 0 | 0.000 | 0.00 |
| 9 | female | 18 | 20 | 19,057 | 0.904 | 37.93 |
| 9 | male | 16 | 75 | 21,984 | 2.106 | 48.00 |
| 10 | female | 5 | 0 | 3,067 | 0.000 | 0.00 |
| 10 | male | 7 | 42 | 11,345 | 2.217 | 58.33 |
| 11 | female | 1 | 0 | 367 | 0.000 | 0.00 |
| 11 | male | 2 | 1 | 943 | 0.646 | 50.00 |
| 12 | female | 1 | 1 | 173 | 5.780 | 100.00 |
| 12 | male | 0 | 0 | 0 | 0.000 | 0.00 |
| Total |  | 46 | 232 | 278,895 | 0.514 | 91.30 |

Table 2 shows that the final data set consists of 46 speakers, 232 instances of DM like, and 278,895 words and that the speakers are distributed rather unevenly across ages groups as there is e.g. only 1 speaker aged 8. The largest sub-cohort with 46 speakers consists of children being recorded at age 5. The highest number of instances of DM like within the sub-cohorts is 95 among male children recorded at age 9. The highest mean value of DM like per 1,000 words in the data amounts to 5.780 which is caused, however, by a single female child using 1 instance of DM like with a conversation encompassing 173 words. In general, it appears as if both the relative frequencies and the percentages of children using DM like seem to increase with age.

Table 3: Summary of the final data set with respect to age and situation type (age groups were collapsed to encompass two years and age cohorts with less than 10 speakers were removed)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| SitType | Age | Speakers | DM like | Words | Mean DM like | Percent |
|  |  | (N) | (N) | (N) | (ptw) | (DM like user) |
| formal | 3-4 | 46 | 3 | 25,338 | 0.159 | 6.12 |
| informal | 3-4 | 46 | 23 | 75,827 | 0.263 | 24.59 |
| formal | 5-6 | 46 | 8 | 29,499 | 0.478 | 13.21 |
| informal | 5-6 | 46 | 17 | 50,770 | 0.318 | 22.41 |
| formal | 7-8 | 41 | 5 | 21,153 | 0.197 | 10.87 |
| informal | 7-8 | 27 | 37 | 19,372 | 1.563 | 46.43 |
| formal | 9-10 | 43 | 60 | 35,086 | 1.101 | 35.59 |
| informal | 9-10 | 32 | 77 | 20,367 | 2.461 | 56.25 |
| Total |  | 46 | 230 | 277,412 | 0.514 | 91.3 |

Table 3 indicates that the highest frequencies of DM like occur during informal conversations and that the percentages of children that use DM like increase with age. It also shows that most of the children were recorded for the first time when they were 3 or 4 years of age.

3.6 Visualization of the final data set

This section displays and summarizes the results of the analysis. Figures 5 and 6 display the distribution of DM like across age groups and genders in terms of percentages and relative frequencies respectively.

|  |  |
| --- | --- |
| C:\03-MyProjects\06LIKEAcquisition\Article\images\likeagesexpct.png  Figure 5: Percentage of children that have used DM like by age (only age groups with more than 5 speakers and who used more than 10,000 words are plotted) | C:\03-MyProjects\06LIKEAcquisition\Article\images\likeagegenderptw.png  Figure 6: DM like by age and gender |

Figure 5 strongly suggests an effect of age on DM like usage as the mean frequency of DM like per speech unit is substantially higher than among older children. In addition, Figure 5 indicates that the use of DM like remains rather stable between 3 or 4 and up to age 9 when its frequency suddenly increases. Figure 6 indicates that male speakers tend to use DM like slightly more than female speakers do.

|  |  |
| --- | --- |
| C:\03-MyProjects\06LIKEAcquisition\Article\images\likeagestyppct.png  Figure 7: Percentage of children that have used DM like by age and situation type | C:\03-MyProjects\06LIKEAcquisition\Article\images\likeagestypptw.png  Figure 8: DM like by age and situation type |

Figure 7 shows that the use of DM like is consistently higher in informal versus formal situation types and that there appears to be a near-linear, monotonic increase in these percentages. It is remarkable, that already with 7 and 8 years of age, nearly half of children in the data use DM like during informal situations. In contrast, only 35.6 percent of children aged 9 and 10 in the data have used DM like during informal situations. It is also remarkable that 24.6 percent of the children aged 3 and 4 have used DM like in informal settings.

|  |  |
| --- | --- |
| C:\03-MyProjects\06LIKEAcquisition\Article\images\likeagemotptw.png  Figure 9: DM like users and non-users by mothers’ frequency of DM like by age | C:\03-MyProjects\06LIKEAcquisition\Article\images\likehvchimotptw.png  Figure 10: DM like by children and mothers against age and situation type |

Figure 9 suggests an effect of mothers’ use of DM like and the likelihood of children to fall into the category of DM like user versus non-users of DM like. During the early stages of acquisition, there does not appear to be an impact of the mothers’ use of DM like on the likelihood that a child uses this marker as the probability of a child to use DM like is almost identical until age 6. From age 6 onwards, the frequency of DM like use by the mother appears to correlate positively with the likelihood of children to be DM like users.

Figure 10 indicates that the frequency of DM like use by children and mothers is very similar until age 6 when a peak in DM like use by children in informal settings is observable. At age 7, the use of DM like by mothers increases very noticeably, especially in informal situation, while children show a more steady and monotonic increase.

3.7 Regression modeling

The study uses a type of multivariate analysis to statistically test if any of the independent variables or interactions between them correlate with a child's usage of DM like in the HSLLD data. More specifically, the primary tool to investigate which factors impact the occurrence of DM like in a child's linguistic output is a mixed-effects binomial logistic regression model.[[5]](#endnote-5) Binomial logistic regression models calculate the likelihood of a binary outcome (child is a user of DM like versus child is not a DM like user) given the independent variables. For instance, a binomial logistic regression model calculates the likelihood of a child being a DM like user given the child is a five-year old male recorded during an informal setting.

To obtain accurate estimates for the size of the effect of a given variable, the model was fit using step-wise step-up (independent variables and their interactions are added consecutively, i.e. the model is build up). The model fitting process is necessary to arrive at a final minimal adequate model, i.e. the best model in the sense that a minimum of predictors explains a maximum of variation. Four criteria were applied to arrive at the most parsimonious model:

1. decrease in the AIC (Akaike information criterion) as this is a measure for parsimony (cf. Field, Miles and Field 2012:263);
2. applying a step-wise step-down procedure analogous to the step-wise step-up method exemplified in Gries (2009:259-263) during which only predictors which lead to a significant improvement in model fit are retained;
3. variance inflation factors (VIFs) – if the VIFs exceeded values greater than 10 (cf. Szmrecsanyi 2006:215), the model was deemed too unstable due to multicollinearity. This means that even though a model might have the lowest AIC it was abandoned if the VIFs were unacceptably high.
4. higher-level interactions were included only if they did not cause cases of complete separation or lead to failures for the model to converge.

The step-wise step-up model fitting procedure arrived at a highly significant final minimal adequate model (cf. Table 7 in the Appendix).

4 Results of the minimal adequate mixed-effects logistic regression model

The final minimal adequate mixed-effects binomial logistic regression model[[6]](#endnote-6) performs significantly better than a base-line model[[7]](#endnote-7) and reports the age of children[[8]](#endnote-8), and the situation type[[9]](#endnote-9) as significant predictors. Although models m6.glmer (added predictor aged:motlisthvptw) and m8.glmer (added predictor sex:styp) both exhibited a decrease in the AIC value and the interactions significantly improved the model fit, the AICs of predictors vastly exceeded a value of 10. Therefore, the models were too unstable to warrant reliable conclusions. The final minimal adequate model did neither confirm that the gender of children affects the probability of a speaker having used DM like[[10]](#endnote-10) nor did it confirm a significant impact of the mother's use of DM like[[11]](#endnote-11) on the probability of a child using this particle in a given speech situation. In addition, the model did not report any significant interactions but including the random effect (random intercepts) was justified as the AIC of the mixed-effects model (184.8) was smaller than the AIC of the fixed-effects model (186.05).

The situation type has the strongest impact on the usage of DM like: during informal situations including meal-time coversations and toy play the probability of a child being a DM like user is 2.71 higher than during formal situation types such as elicited reports, reading, and letter writing (the baseline or reference category). Age also substantially and significantly affects the likelihood of DM like usage: the probability of speaker becoming a DM like user increases by a factor of 1.61 per year.

5 Discussion and Outlook

The current analysis of the use of DM like of 46 children from New England between the ages of 3 and a half and 12 years during different situation types of has shown that DM like is already present in the speech of 3- and 4-year olds and that the use of DM like is acquired together with its extra-linguistic constraints as its use is substantially more common in informal compared to formal situation types. Thus, the data used in this study allowed to investigate the acquisition of variation in the use of an innovative feature of English during relatively early stages of L1-acquisition.

The results of the mixed-effects model show that DM like is used particularly often during informal situations such as family meal time conversations, i.e. a situation type during which investigators were not present, or toy play. This strongly suggests that children are not only aware of extra-linguistic constraints at a very young age but that they are aware of stylistic adequateness of variants and variation governed by the settings surrounding speech events. In addition, the fact that including random intercepts significantly increased the model fit shows that there are significant differences between children and the age when they acquire DM like. In other words, while there are common pathways and recurrent patterns during L1-acquisition, children differ with respect to when they reach certain stages. Although this seems trivial, it shows that language acquisition is speaker-specific and that individuals differ in their acquisition of linguistic features both with respect to the age of acquisition and constraints on its use once they have acquired a linguistic feature. Astonishingly, if the percentage of children who have used DM like in any of their recordings is calculated, 91.3 percent of children have in fact used DM like (cf. Table 3). Furthermore, the probability of a child having used DM like in informal settings compared to formal speech situations is consistent across age groups and apparent even among very young children (cf, Figure 7). This finding is intriguing as it suggests that children are highly sensitive to the situation type (and thus formality or register as well as stylistic variation) and that they adapt their language accordingly. To elaborate, depending on whether the children are in an in-group context or in a situation in which non-in-group interlocutors are present, they appear to constraint their linguistic output and hence use less non-standard features. It is particularly revealing that the situation type has an even stronger impact than the age of children which suggests that their awareness of register variation and thus stylistic constraints on the use of certain variants is acquired very early. This may in fact explain why Miller & Weinert (1995) hypothesized that DM like is acquired after an age of approximately 10 – this hypothesis was supported by the extremely low frequencies of DM like in the speech of northern Irish children between the ages of 9 and 12 in Schweinberger (2013). Another possible explanation for the late-acquisition-hypothesis could be that while DM like is acquired very early on, the frequency of use increases substantially at age 9 for boys and 10 for girls (cf. Figure 6). Thus, the early acquisition of DM like could be overlooked due to its low frequency among children younger than 9 or 10. As argued above, this could in addition to register constraints which were discussed above lead investigators to erroneously conclude that DM like is acquired rather late.

The age of children is also a highly significant factor as the probability of a child becoming a DM like user increases by a factor of 1.61 per year. The percentages of DM like users among children appear to remain stable between the ages of 3 and 4 to about 9 when its frequency increases notably (cf. Figure 5). This finding could be explained in at least two different ways: first, the increase in frequency is associated with the impact of their peer group on the children’s linguistic output. This may even be expected given the shift in focus from caregivers to peers as linguistic role models during early adolescence (cf. Eckert, 1999). However, Figure 9 suggested the opposite, i.e. that a higher frequency of DM like in the mothers’ input increases the likelihood of a child using this particle. An alternative would be that the increase is related to a functional expansion of DM like use and the frequency of DM like increases because children use this particle in innovative ways which allows it to intrude contexts from which it was formerly barred. This explanations is supported by analyses of other multifunctional discourse features which also exhibited a substantial increase after children began attending school (cf. Tomasello, 2003:269).

Whatever the exact cause for this increase in frequency may be, the results show that a substantial percentage of children as young as 3 or 4 years of age have acquired DM like which is in accordance with research founding that the period between 3 and 4 years of age is crucial for acquiring dialectal norms and constraints on variation within the speech community (Roberts & Labov, 1995:110). In addition, the findings support research according to which children are not passively acquiring variants but rather actively partake in ongoing language change as previously suggested, for instance, by Levey (2006), Roberts & Labov (1995), and Roberts (2008).

Furthermore, the present analysis casts a shadow of doubt on claims according to which DM like is acquired rather late, i.e. at approximately age 10 and outright contradicts an early model of dialect acquisition proposed by Labov (1964) which assumes that children remain monodialectal until age 10 and only subsequently acquire extra-linguistically governed (stylistic) variation (cf. Nardy, Chevrot, and Barbu, 2013:258-260). The data support the hypothesis that structures that are prominent in un-monitored face-to-face conversations are structures that are acquired relatively early on and are thus – from a usage-based model of language acquisition – deeply entrenched which is compatible with the model described by (Ochs, 1979). Also, the present results lend support to modern usage-based models of language acquisition or construction grammar approaches to language acquisition (cf. e.g. Hilpert, 2014:155-178) that emphasize the importance of the linguistic input especially during early stages of language acquisition and which take a usage-based approach towards language acquisition (Tomasello, 2003).

While the absence of a gender difference is in line which the hypothesis that although dialect features are acquired early on, socially significant variation emerges during adolescence (Labov, 1964), the absence of a direct impact of the mothers’ use of DM like is unexpected. However, Figure 9 suggested that there is an interaction between the frequency with which mothers use DM like and the age at which children acquire DM like. According to Figure 9, the use of DM like by mothers appears to impact on the likelihood of children to become DM like users as the higher the frequency of the mother, the more likely it is that a child uses DM like. While the graphical display is intriguing as both variationist theories on language transmission (Labov, 1994) and usage-based models of language acquisition (Tomasello, 2003; cf. also Saxton, 2010) predict that the caregiver's language is the primary input affecting the child's linguistic output, the regression model did not confirm this interaction as being a significant predictor. A likely explanation for the lack of a significant interaction between the mothers’ use of DM like and the age at which children acquire DM like may be the relatively small data set. Although the HSLLD is a very interesting data set, it only contains 46 children that have been present during all home visits and situation types. Whatever the case may be, the visualization in Figure 9 can only lend tentative support to studies that propose that as children mature, their focus shifts from adults as linguistic role models to their peers (Eckert, 1999; Labov, 1972).

One of the issues that remain unsolved relates to the functional employment of DM like which was proposed above as a possible cause of the notable increase in the frequency of DM like once children turn nine. DM like can be used to fulfill various functions, e.g. to expand the truth conditions of sentences (cf. Siegel, 2002), to express vagueness or a difference between what is said and what is meant (Schourup 1982), to highlight new information (Miller & Weinert, 1995; Underhill, 1988), to buy processing time (Schweinberger, 2014). The present study has differentiated functionally distinct variants. It appears to be a reasonable hypothesis that distinct functions are acquired at different stages and that the increase in the use of DM like may not necessarily be attributed to a pure increase in frequency but rather a functional expansion of its use. More fine-grained analyses are required to address this remaining issue.

Another issue which could not be addressed here but that warrants closer inspection relates to what extend the use of DM like by children during L1 acquisition is affected by identify construction, i.e. to what extend children use DM like is utilized as an identify marker or marker of group membership among children. Unfortunately, this intriguing question would in any case be difficult to study due to the illusive and multivariate nature of identity construction. However, the stage in life at which children start to express and construct their identity linguistically is a highly relevant topic of variationist research. On a related note, further analysis is needed to assess the shift from caregivers to peers as linguistic role models more reliably.

7 Endnotes

8 References

Andersen, Gisle. (1997). They like wanna see like how we talk and all that. The use of like as a discourse marker in London teenage speech. In M. Ljung (Ed.), Corpus-based studies in English. Amsterdam: Rodopi. 37-48.

Andersen, Gisle. (1998). The pragmatic marker like from a relevance-theoretic perspective. In A. H. Jucker & Y. Ziv (Eds.), Discourse markers: descriptions and theory. Amsterdam & New York: John Benjamins Publishing Company. 147-170.

Andersen, Gisle. (2000). The role of the pragmatic marker like in utterance interpretation. In G. Andersen & T. Fretheim (Eds.), Pragmatic Markers and Propositional Attitude. Amsterdam & Phildelphia, PA: John Benjamins Publishing Company. 17-38.

Andersen, Gisle. (2001). Pragmatic markers and sociolinguistic variation: A relevance-theoretic approach to the language of adolescents, Volume 84. Amsterdam & Phildelphia, PA: John Benjamins Publishing Company.

Anderson, Anne. H., Clark, Aileen & Mullin, James. (1991). Introducing information in dialogues: forms of introduction chosen by young speakers and the responses elicited from young listeners. Journal of Child Language 18(3): 663-687.

Baayen, R. Harald. (2008). Analyzing linguistic data. A practical introduction to statistics using R. Cambridge: Cambridge University Press.

Barbu, Stéphanie, Nardy, Aurélie, Chervot, Jean-Pierre & Juhel, Jacques. (2013). Language evaluation and use during early childhood: Adhesion to social norms or integration of environmental regularities?. Linguistics 51(2): 381-411.

Behrens, Heike. (2009). Usage-based and emergentist approaches to language acquisition. Linguistics 47(2): 383–411.

Bybee, Joan. (2007). Frequency of Use and the Organization of Language. Oxford: Oxford University Press.

Bybee, Joan. (2010). Language, Usage, and Cognition. Cambridge: Cambridge University Press.

Dailey-O’Cain, Jennifer. (2000). The sociolinguistic distribution of and attitudes toward focuser like and quotative like. Journal of Sociolinguistics 4(1): 60-80.

CHILDES Project. (unknown). CHILDES data base manuals. North American English Corpora. url: http://childes.psy.cmu.edu/manuals/02english-na.pdf (accessed December 8th, 2015).

D’Arcy, Alexandra F. (2005). Like: Syntax and development. Doctoral dissertation, University of Toronto.

D’Arcy, Alexandra F. (2007). Like and language ideology: Disentangling fact from fiction. American Speech 82(4): 386-419.

Eckert, Penelope. (1999). Linguistic Variation as Social Practice: The Linguistic Construction of Identity in Belten High. Oxford: Blackwell Publishers.

Field, Andy, Miles, Jeremy & Field, Zoe. (2012). Discovering statistics using R. Sage.

Fuller, Janet M. (2003). Use of the discourse marker like in interviews. Journal of Sociolinguistics 7(3): 365-377.

Grant, William & Dixon, James M. (1921). Manual of modern Scots. Cambridge: Cambridge University Press.

Hilpert, Martin. (2014). Construction Grammar and its application to English. Edinburgh: Edinburgh University Press.

Ibbotson, Paul. (2013). The scope of usage-based theory. Frontiers in Psychology 4: 1-15.

Jespersen, Otto. (1954). A Modern English Grammar on Historical Principles: Part 6 - Morphology. London: Allen & Unwin.

Kirk, John. (1992). The Northern Ireland Transcribed Corpus of Speech. In G. Leitner (Ed.), New directions in English language corpora. methodology, results, software developments. Berlin & New York: Mouton de Gruyter. 65-73.

Labov, William. (1964). Stages in the acquisition of Standard English. In R. Shuy (Ed.), Social Dialects and Language Learning. Champaign, IL: National Council of Teachers of English. 77-103.

Labov, William. (1972). Language in the inner City: Studies in the Black English Vernacular, Volume 3. Philadelphia, PA: University of Pennsylvania Press.

Labov, William. (1994). Principles of Language Change: Internal Factors, Volume 1. Oxford: Blackwell.

Levey, Steven. (2006). The sociolinguistic distribution of discourse marker like in preadolescent speech. Multilingua-Journal of Cross-Cultural and Interlanguage Communication 25(4): 413-441.

Lieven, Elena & Tomasello, Michael. (2008). Children’s first language acquisition from a usage-based perspective. In P. Robinson & N. C. Ellis (Eds.), Handbook of Cognitive Linguistics and Second Language Acqusition. Routeledge: New York & London. 168-196.

Macaulay, Ronald, K. S. (1977). Language, social class, and education: A Glasgow study. Edinburgh: Edinburgh University Press.

MacWhinney, Brian. (2000). The CHILDES project: The database, Volume 2. Psychology Press.

Meehan, Teresa. (1991). It's like, ʻwhat's happening in the evolution of like?ʼ: A theory of grammaticalization. Kansas Working Papers in Linguistics 16: 37-51.

Miller, Jim & Weinert, Regina. (1995). The function of like in dialogue. Journal of Pragmatics 23 (4): 365-393.

Miller, Jim & Weinert, Regina. (1998). Spontaneous Spoken Language: Syntax and Discourse. Oxford: Clarendon Press.

Nardy, Aurélie, Chervot, Jean-Pierre & Barbu, Stéphanie. (2013). The acquisition of sociolinguistic variation: Looking back and thinking ahead. Linguistics 51(2): 255-284.

Ochs, Elinor. (1979). Planned and unplanned discourse in discourse and syntax. In T. Givon (Ed.), Syntax and Semantics. Vol. 12: Discourse and syntax. New York: Academic Press. 51-80.

Partridge, Eric. (1984). A Dictionary of Slang and unconventional English. London: Routledge & Kegan Paul.

Pichler, Heike & Levey, Steven. (2010). Variability in the co-occurrence of discourse features. University of Reading Language Studies Working Papers 2: 17-27.

Roberts, Julie. (1996). Acquisition of Variable Rules: (-t,d) deletion and (ing) production in preschool children. Institute for Research in Cognitive Science (IRCS) Report 96-09. Philadelphia: University of Pennsylvania.

Roberts, Julie. (1997). Acquisition of variable rules: a study of (-t, d) deletion in preschool children. Journal of Child Language 24 (2): 351-372.

Roberts, Julie. (2008). Child language variation. In J. K. Chambers, P. Trudgill & N. Schilling-Estes (Eds.), The Handbook of Language Variation and Change. Oxford: Blackwell. 252-263.

Roberts, Julie & Labov, William. (1995). Learning to talk Philadelphian: Acquisition of short a by preschool children. Language Variation and Change 7 (1): 101-112.

Romaine, Suzanne. (1978). Postvocalic /r/ in Scottish English: Sound change in Progress. In P. Trudgill (Ed.), Sociolinguistic Patterns in British English. Baltimore, MD: University Park Press. 144-157.

Romaine, Suzanne & Lange, Deborah. (1991). The use of like as a marker of reported speech and thought: A case of grammaticalization in progress. American Speech 66(3): 227-279.

Saxton, Matthew. (2010). Child Language: acquisition and development. Los Angeles: Sage.

Schmid, Hans-Jörg. (2007). Entrenchment, salience and basic levels. In: Dirk Geeraerts and Hubert Cuyckens (Eds.), The Oxford Handbook of Cognitive Linguistics. Oxford University Press: Oxford. 117-138

Schmid, Hans-Jörg. (2010). Does frequency in text really instantiate entrenchment in the cognitive system?. In D. Glynn & K. Fischer (Eds.), Quantitative methods in cognitive semantics: Corpus-driven approaches. Walter de Gruyter: Berlin. 101-133.

Schourup, Lawrence C. (1982). Common discourse particles in English conversation. Doctoral dissertation, Ohio State University.

Schweinberger, Martin. (2013). A sociolinguistics analysis of discourse marker like in Northern Ireland. A look behind the scenes of quantitative reasoning. In M. Bieswanger & A. Koll-Stobbe (Eds.), New Approaches to the Analysis of Linguistic Variability. Frankfurt a.M.: Peter Lang. 13-39.

Schweinberger, Martin. (2014). The discourse marker LIKE: a corpus-based analysis of selected varieties of English. Doctoral dissertation, University of Hamburg.

Siegel, Muffy E. A. (2002). Like: The discourse particle and semantics. Journal of Semantics 19 (1): 35-71.

Simpson, John A., Weiner, Edmund S. & et al. (1989). The Oxford English Dictionary, Volume 8. Oxford: Clarendon Press Oxford.

Smith, Jennifer, Durham, Mercedes & Fortune, Liane. (2007). Community, caregiver and child in the acquisition of variation in a Scottish dialect. Language Variation and Change 19(1): 63-99.

Smith, Jennifer, Durham, Mercedes & Richards, Hazel. (2013). The social and linguistic in the acquisition of sociolinguistic norms: Caregivers, children, and variation. Linguistics 51(2): 285-324.

Szmrecsanyi, Benedikt. (2006). Morphosyntactic Persistence in Spoken English. A corpus study. Berlin & New York: Mouton de Gruyter.

Tagliamonte, Sali. (2005). So who? like how? just what?: Discourse markers in the conversations of young Canadians. Journal of Pragmatics 37(11): 1896-1915.

Tomasello, Michael. (2003). Constructing a language: A usage-based theory of language acquisition. Boston: Harvard University Press.

Tomasello, Michael. (2012). The usage-based theory of language acquisition. In Edith L. Bavin (Ed.), Cambridge Handbook of Child Language. Cambridge: Cambridge University Press. 69-88.

Underhill, Robert. (1988). Like is, like, focus. American Speech 63(3): 234-246.

Wright, Joseph. (1902). The English Dialect Dictionary, Volume 4. London: H. Frowde.

Wright, Thomas. (1857). Dictionary of obsolete and provincial English, Volume 2. London: Henry G. Bohn.

9 Appendix

Table 4: Overview of age of children by home visit

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Age | Homevisit 1 | Homevisit 2 | Homevisit 3 | Homevisit 5 | Homevisit 7 |
| 3 | 86 | - | - | - | - |
| 4 | 76 | 136 | 5 | - | - |
| 5 | - | 24 | 209 | 1 | - |
| 6 | - | 1 | 5 | 11 | - |
| 7 | - | - | - | 107 | 3 |
| 8 | - | - | - | - | 3 |
| 9 | - | - | - | - | 86 |
| 10 | - | - | - | 1 | 28 |
| 11 | - | - | - | - | 5 |
| 12 | - | - | - | - | 1 |

Table 5: Overview of previous research on the use of DM like

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Study | Miller & Weinert (1995) | | | Andersen (2001) | D'Arcy (2005) | | Levey (2006) | | | | | Schweinberger (2013) | | | |
| Region | Glasgow | | | London | Toronto | | London | | | | | Northern Ireland | | | |
| Date | 1988 | | | 1993 | 2002-2004 | | 2000-2004 | | | | | 1973-1980 | | | |
| Words (N) | NA | | | 106,798 | NA | | 70,134 | | | | |  | | 50,915 | |
| Speakers (N) | 66 | 52 | 52 | NA | 5 | 5 | 21 | | 27 | | | 19 | | 17 | |
| Gender | NA | NA | NA | NA |  |  | male | female | | male | female | | male | | female | |
| Age | 8 | 10 | 13 | 10-13 | 10-12 | 10-12 | 7-8 | 10-11 | | 7-8 | 10-11 | | 9-12 | | 9-12 | |
| Absolute Freq. (N) | 2 | 4 | 23 | 244 | NA | NA | 119 | 78 | | 85 | 137 | | 15 | | 16 | |
| Relative Freq. (ptw) | NA | NA | NA | 2.53 | NA | NA | 6.75 | 4.97 | | 5.03 | 6.89 | | 0.43 | | 0.34 | |

Table 6: Results of the final minimal adequate mixed-effects binominal logistic regression model

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Group(s) | Variance | Std. Dev. |  |  | L.R. χ2 | DF | Pr | Significance |
| Random Effect(s) | childid | 0.57 | 0.75 |  |  | 3.3 | 1 | 0.0694 | p < .10(\*) |
| Fixed Effect(s) | Estimate | VIF | OddsRatio | CI(2.5%) | CI(97.5%) | Std. Error | z value | Pr(>|z|) | Significance |
| (Intercept) | -5.44 |  | 0 | 0 | 0.03 | 0.96 | -5.68 | 0 | p < .001\*\*\* |
| aged | 0.48 | 1.03 | 1.61 | 1.31 | 1.98 | 0.11 | 4.54 | 0 | p < .001\*\*\* |
| stypinformal | 1 | 1.03 | 2.72 | 0.98 | 7.6 | 0.52 | 1.91 | 0.0555 | p < .10(\*) |
| Model statistics | |  |  |  |  |  |  |  | Value |
| Number of Groups | |  |  |  |  |  |  |  | 31 |
| Number of cases in model | |  |  |  |  |  |  |  | 189 |
| Observed misses | |  |  |  |  |  |  |  | 153 |
| Observed successes | |  |  |  |  |  |  |  | 36 |
| Residual deviance | |  |  |  |  |  |  |  | 154.57 |
| R2 (Nagelkerke) | |  |  |  |  |  |  |  | 0.232 |
| R2 (Hosmer & Lemeshow) | | |  |  |  |  |  |  | 0.16 |
| R2 (Cox & Snell) | |  |  |  |  |  |  |  | 0.144 |
| C |  |  |  |  |  |  |  |  | 0.863 |
| Somers‘ Dxy | |  |  |  |  |  |  |  | 0.726 |
| AIC |  |  |  |  |  |  |  |  | 162.57 |
| BIC |  |  |  |  |  |  |  |  | 175.53 |
| Prediction accuracy | |  |  |  |  |  |  |  | 84.13% |
| Model Likelihood Ratio Test | | |  |  |  | L.R. χ2: | 29.49 | DF: 3 | p<.001\*\*\* |

Table 7: Results of the model fitting process (step-wise step-up): mixed-effects binominal logistic regression model

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | Term Added | Compared to... | DF | AIC | BIC | LogLikelihood | Residual Deviance | χ2 | χ2DF | p-value | Significance |
| m1.glmer | aged | m0.glmer | 3 | 164.66 | 174.38 | -79.33 | 158.66 | 22.1 | 1 | 0 | p<.001\*\*\* |
| m2.glmer | sex | m1.glmer | 4 | 165.67 | 178.64 | -78.84 | 157.67 | 0.99 | 1 | 0.32014 | n.s. |
| m3.glmer | motlisthvptw | m1.glmer | 4 | 166.47 | 179.43 | -79.23 | 158.47 | 0.19 | 1 | 0.66063 | n.s. |
| m4.glmer | styp | m1.glmer | 4 | 162.57 | 175.53 | -77.28 | 154.57 | 4.09 | 1 | 0.04309 | p<.05\* |
| m5.glmer | aged:sex+sex | m4.glmer | 6 | 165.12 | 184.57 | -76.56 | 153.12 | 1.45 | 2 | 0.48394 | n.s. |
| m6.glmer | aged:motlisthvptw+motlisthvptw | m4.glmer | 6 | 159.63 | 179.08 | -73.82 | 147.63 | 6.94 | 2 | 0.03116 | p<.05\* |
| m7.glmer | aged:styp | m4.glmer | 5 | 164.09 | 180.3 | -77.05 | 154.09 | 0.48 | 1 | 0.4896 | n.s. |
| m8.glmer | sex+sex:styp | m4.glmer | 6 | 158.03 | 177.48 | -73.01 | 146.03 | 8.54 | 2 | 0.01399 | p<.05\* |

1. According to the respective manual, the HSLLD corpus is said to feature 100 children and covers an age range of two to six year-olds. However, a closer inspection showed that the HSLLD corpus features only 85 children covering an age range of three to twelve year-olds. [↑](#endnote-ref-1)
2. An overview of the exact age that children were during the home visits in provided in Table 4 in the Appendix. [↑](#endnote-ref-2)
3. For a detailed description of the making-up of family income, the reader is referred to the relevant section in the CHILDES manual (CHILDES Project, unknown). [↑](#endnote-ref-3)
4. For graphical representations and summaries of the data set, this numeric variable was re-coded as an order factor, i.e. a categorical variable. [↑](#endnote-ref-4)
5. In a first attempt, a linear mixed-effects model was fitted to the data with the relative frequency of DM like use of children as the dependent variable to the data. Unfortunately, this model violated various conditions such as heteroscedasticity, kurtosis, and skewness (cf. Field, Miles & Field, 2012 for a discussion of the effects and nature of violations of model conditions) and had thus to be abandoned. Using a linear mixed-effects model would, however, been preferable as a numeric dependent variable would have been more informative than a nominal dependent variable which is the model reported here. [↑](#endnote-ref-5)
6. For a more extensive overview of the results cf. Table 6. [↑](#endnote-ref-6)
7. Model Likelihood Ratio Test: L.R. χ2: 29.49, DF: 3, Significance: p-value <.001\*\*\*. [↑](#endnote-ref-7)
8. χ2: 22.096, DF: 1, Significance: p-value <.001\*\*\*. [↑](#endnote-ref-8)
9. χ2: 4.0917, DF: 1, Significance: p-value <.05\* (.043). [↑](#endnote-ref-9)
10. χ2: 0.9884, DF: 1, p-value = 0.32 (cf. also Table 7). [↑](#endnote-ref-10)
11. χ2: 0.1928, DF: 1, p-value = 0.66 (cf. also Table 7). [↑](#endnote-ref-11)