L1-acquisition of the discourse marker like

The L1-acquisition of the discourse marker like

Martin Schweinberger
University of Hamburg
Department of English and American Studies
Von-Melle-Park 6
D - 20146 Hamburg
Germany
Fon: +49 -(0)40-42838-3968
E-mail: martin.schweinberger@uni-hamburg.de

Tayo Takada
Max Planck Institute for Psycholinguistics
PO Box 310
6500 AH Nijmegen
The Netherlands

Abstract

This study takes a quantitative, variationist approach to investigate the acquisition of the discourse marker like (DM like) based on the Home-School Study of Language and Literacy Development (HSLLD) component of the Child Language Data Exchange System (CHILDES). The results show that DM like is already present in the speech of 4-year old children and that the likelihood of children to use DM like correlates strongly with the situational setting. The results thus support previous research claiming that children may already acquire systematic variation between the ages of 3 and 4 while they contrast with claims that DM like is acquired after the age of 10 which is particularly interesting from a theoretical perspective as the late acquisition of DM like has been brought forward against a model of discourse production according to which discourse structures which are frequently used in unplanned discourse are acquired particularly early during L1 acquisition.
L1-acquisition of the discourse marker like

1 Introduction

While there has been a growing body of research in variationist sociolinguistics on pre-adolescents, issues relating to the acquisition of variation during very early stages of L1-acquisition still remain unsolved mainly due to methodological problems (e.g. small sample sizes, the use diary entries rather than natural language data, or lack of sufficiently large data sets) and theoretical difficulties such as distinguishing between developmentally and socially motivated variation (cf. Roberts, 2008). In addition, while most studies which have focused on linguistic variation during L1 acquisition have addressed phonological variation such as (t,d) deletion (Roberts, 1997), (ing) realization (Roberts, 1996), the production of word-final (r) (Romaine 1978) or short (a) (Roberts & Labov, 1995), few studies have addressed variable use of lexical items functioning as pragmatic markers. This is unfortunate as studies suggest that children actively partake in ongoing change and may indeed accelerate changes in progress (cf. e.g. Levey, 2006; Roberts, 2008). Another interesting issue which has so far received only marginal attention relates to the import of the input by primary caregivers versus peers. Eckert (1999) suggests that when children enter secondary school, their focus shifts from adults to peers as primary locus of linguistic input. The present analysis addresses the above mentioned issues by investigating the use of DM like during early stages of L1 acquisition.

Although the discourse marker like\(^1\) (cf. (1)), i.e. a syntactically optional and semantically bleached use of <like>\(^2\), has been depicted as a traditional, dialectal, non-standard or even a vulgar feature of English (e.g. Grant & Dixon, 1921; Jespersen, 1954; Partridge, 1984; Wright, 1857; Wright, 1902;), DM like has successfully conquered the English speaking world and is among the best studied discourse markers in English (e.g. Andersen, 1997, 1998, 2000, 2001; Dailey-O’Cain, 2000; D’Arcy, 2005, 2007; Fuller 2003;
L1-acquisition of the discourse marker *like*


(1)

a. *CHI: an(d) *like (.) with the ham juice in your mouth *like you get *like thirsty . (HV2/MT/remmt2)

b. *MOT: it looks like maybe he's *like maybe you know two (.) maybe three but I don't even think he looks three . (HV2/MT/inamt2)

c. *CHI: and I can (.) write it *like &um (.) this side ? (HV5/LW/meglw5)

While previous research has addressed issues relating to DM *like* use among (pre)adolescents (D’Arcy, 2005; Levey, 2006; Pichler & Levey, 2010) and even the acquisition of DM *like* (Miller & Weinert, 1995, 1998; Schweinberger, 2013), none of these studies have empirically investigated the earliest stages of language acquisition with respect to this pragmatic marker. This is a crucial gap in our knowledge for at least two reasons: firstly, recent studies (e.g. Roberts, 1997; Roberts & Labov, 1995) have found that children acquire patterns of variation and the respective constraints already between the ages of 3 and 4 which poses the question of whether this also holds true for DM *like*. Secondly, previous research has claimed that DM *like* is acquired rather late, i.e. after an age of 10 (Miller & Weinert, 1995). However, if DM *like* is acquired that late, then this would not only be at odds with Roberts & Labov (1995) but it would also confute predictions that can be derived from a model of discourse production proposed by Ochs (1979). According to Ochs (1979:53) speakers rely on morphosyntactic and discourse skills acquired between the age of 3 and 4 when engaging in unplanned discourse. However, Miller & Weinret (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 10 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-
L1-acquisition of the discourse marker **like**
adolescent speech use data of speakers unfit to investigate the use of DM **like** during early stages of L1-acquisition.

The present study aims to close this gap in our knowledge about the acquisition of 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).

The following section provides an overview of the previous research on the acquisition of DM **like** and its use in pre-adolescent speech. Section 3 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. Finally, section 6 offers a conclusion and an outlook.

2 Previous research

A typically developing child produces its first word around age one, understands about 50 words with 18 months, and has a vocabulary of some 6,000 words by the age of 5 (Saxon, 2010:7-8, 17). It seems natural to assume that DM **like**, a feature ever so prominent in adolescent speech (Andersen, 2001:226), should be part of the vocabulary of a significant proportion of 5-year-olds. Variationist research (Roberts, 1997; Roberts & Labov, 1995) lends support to this view as even children aged 3 and 4 are able to acquire phonological variants along with the respective constraints on stable variation. To elaborate, Roberts & Labov (1995) found that children from Philadelphia aged 3 and 4 accurately acquired phonological variants, short **a**, along with the respective constraints and concluded that even very young speakers take part in ongoing language change. This finding can be substantiated theoretically by a model of discourse production (Ochs, 1979) which argues that, due to the heavy processing load in the on-line speech production process, adult speakers tend to rely more
L1-acquisition of the discourse marker *like*

heavily on morphosyntactic and discourse skills acquired in the first 3 or 4 years of life when engaging in unplanned discourse (Ochs, 1979:53). Given that DM *like* in most prominent in unplanned discourse, this implies that we would expect DM *like* to be acquired relatively early in life.

However, Miller & Weinert (1995:366), who investigated the use of DM *like* in pre-adolescent speech, argue that it is learned relatively late by native-speaker learners. In fact, Miller & Weinert (1995:366) explicitly refer to the model proposed by Ochs (1979) and conclude that it is incorrect with respect to DM *like* and propose that DM *like* is acquired after the age of 10 not because of an intrinsic difficulty a child experiences with this construction but because discourse management skills are acquired rather late (Miller & Weinert, 1995:380). It is important to note, however, that Miller & Weinert (1995) focused on the functions of DM *like* in Scottish English based speakers whose age ranged between 8 and 30 years of age. The data was taken from two different sources which represented spontaneous spoken language recorded during an artificial map task and spontaneous conversations. The *Here Map Task Corpus* (Anderson et al., 1991) contains data of 170 speakers and represents the speech of children between the ages 7 and 8, 9 and 10, and 12 and 13 with roughly balanced distributions across sex who were recorded during an experimental map-task (Anderson et al., 1991:668). Although their primary aim was to investigate the discourse-pragmatic functionality of DM *like* in spoken discourse Miller & Weinert (1995) found that DM *like* is substantially less frequent in the speech of children below an age of 13 compared to the speech of children aged between 7 and 10. This finding led them to conclude that DM *like* is acquired after an age of 10. While Miller & Weinert (1995) convincingly argue that DM *like* is a multifunctional discourse element and not merely a meaningless interjection (as proposed e.g. by Simpson, 1989), their claim concerning the age of acquisition is still up for debate.
In a subsequent study, Levey (2006) analyzed the speech of pre-adolescents and challenged the hypotheses proposed by Miller & Weinert (1995) that DM *like* is not acquired until after the age of 10. His findings show that DM *like* is encountered frequently below an age of 10 (Levey, 2006:415). However, Levey's data consist of speakers between the ages of 7 and 11 and - while being able to falsify Miller & Weinert’s (1995) claim - the data are unfit to test at which age DM *like* is actually acquired. Again, it is important to note that the aim of his study was to investigate the role of pre-adolescents in cases of language change rather than the acquisition of DM *like*. As a result, Levey (2006) concludes that his findings call for a more fine-grained analysis of the function of pre-adolescents during language change.

Andersen (2001) investigated the use of DM *like* in London teenager speech based on The Bergen Corpus Of London Teenage Language (COLT) and finds that DM *like* is a frequent element in the speech of 10 to 13-year-olds with a relative frequency of 2.53 instances per 1,000 words (Andersen, 2001:289). Similar to Levey (2006), his study did not aim to investigate the acquisition of DM *like* but to analyze its discourse-pragmatic functions and its social profile with respect to sociolinguistic variables (age, gender, social class, ethnicity, and location/neighborhood) in order to determine which age cohort is responsible for its spread throughout the London speech community.

D’Arcy (2005) conducted a fine-grained variationist analysis of DM *like* based on the Toronto English Archive which represents 97 speakers aged 10 to 80+ years (D’Arcy, 2005:25). Her study aimed to investigate the intra-linguistic (syntactic) constraints of DM *like* and - more importantly - the social stratification of these constraints and the frequency of distinct variants of DM *like*. Her analysis shows that both the constraints and DM *like* and the frequencies of its functional variants are neatly stratified. Although it is not her focus, her data clearly show that DM *like* is a common discourse element in the speech of 10-year-olds (cf. e.g. D’Arcy, 2005:186).
Schweinberger (2013) analyzed the use of DM like in Northern Irish English based on the Northern Ireland Transcribed Corpus of Speech (NITCS) (Kirk, 1992) which represents speakers between the ages of 9 and 12, 35 and 45, and 65 and 75. His results show that although DM like is used by speakers between the ages of 9 and 12, it is used with a substantially lower frequency compared to the two older cohorts (Schweinberger, 2013:22) which he considered to support the findings of Miller & Weinert (1995) who also investigated a non-standard Celtic-English variety.

Despite their methodological diversity, one feature shared by the studies that have included pre-adolescent speakers is that their analyses rely on data representing speakers who are too old to investigate early stages of L1-acquisition. The youngest age cohort is used in Levey’s (2006) study and comprises speakers between the ages of 7 and 8. The high relative frequencies of the DM like in Levey (2006) warrants a closer look at even younger children. In this respect, the present study addresses the following research questions:

1. At which age is the DM like acquired?
2. Is DM like present in the speech of children aged 3 to 4?
3. Which factors correlate with the use of the 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?

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.
L1-acquisition of the discourse marker like

3.1 Corpus selection

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 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) it represents data collected from the same children over a significant time span and thus allows to track the linguistic development of children, and (vi) it contains speech produced during divers tasks such as book reading, toy playing, etc.

The HSLLD data have been collected between 1988 and 1996 and contain 383,585 words (counting only children’s contributions). 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. 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)
L1-acquisition of the discourse marker like

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 more coherent and quantitatively similar situation types in the final data set (Figure 2).

![Figure 1 Absolute number and percentages of words by situation type in the HSLLD corpus](image1)

![Figure 2: Absolute number and percentages of words by situation type in the recoded data set](image2)

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. Nearly all children are raised in low-income families 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
3.1 L1-acquisition of the discourse marker like
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. Then, the table in which each row represented one instance of DM like was tabulated so that each row would now represent the number of DM like per child per situation type and per visit. In a next step, the relative frequency of DM like by the mothers’ irrespective of visit or situation type were calculated and added to the data table. Also, 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 addition, the relative frequency of DM like was calculated for each row and added to the data to enable visualizing the data. Finally, another variable was created (likecmsgm) which contained information on whether a child had used DM like in the same situation type during a previous recording. If this was the case then this child would be coded as a DM like user for all later recordings of this situation type regardless of whether the child actually used DM like in the later recording or not. To exemplify, if a child had used DM like during the situation
L1-acquisition of the discourse marker *like*
type *meal time* at age 5, the child would be coded for all subsequent *meal time* situations as a
*DM like* user even though it may have not used *DM like* in later recordings of meal time
conversations. The reasoning here is that a given child would only need to show one instance
of *DM like* to be regarded as having successfully acquired *DM like* in a given situation type
while it would be unlikely that the child would unlearn the use of *DM like* once it had learned
its appropriate use. Also, it would have been unfavorable to classify a child as having not
acquired *DM like* in a later recording although it had used *DM like* before given that the later
recording consisted only of relatively little output. 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*.

3.3.1 Non-discourse marker uses

Instances of *<like>* which represented the use of *<like>* as a transitive verb meaning *to fancy*
as in (2), the use of *<like>* as a comparative preposition (cognate with modern German
*gleichen*) which takes a nominal complement as in (3a) or a sentential complement as in (3b)
were coded as non-discourse marker uses of *<like>* and accordingly removed from the
analysis.

\[(2)*CHI: <nuhuh some> [<] boys like ballet and some girls like karate .\]

(HV2/MT/chamt2)
L1-acquisition of the discourse marker *like*

(3) a. *CHI: *<I thought caterpillars> [<] were *like* (. ) humans . (HV7/LW/geolw7)

b. *CHI: *(. ) *like* I do a flip on my bed right ?

*MOT: *(. ) yeah like you shouldn't right ? (HV7/MT/jermt7)

Similarly, general extenders including *<like>* (adjunctive and stuff like that, disjunctive or something like that, and exclusive nothing like that) were also excluded from the analysis.

(4) *CHI: *(. ) and &um we have_ to bring in two dollars to make things like necklaces and stuff like that . (HV7/MT/trimt7)

(5) *CHI: *(. ) <why did> [/] (. ) why [//] &h how can (. ) he find (. ) the fruits or some (. ) and stuff (. ) <with &um> [//] (. ) and (. ) &um without &um like planting them on a tree or something like that ? (HV7/LW/geolw7)

In certain contexts, *<like>* may also function as a conjunction akin to as if. Such instances are also not considered instances of DM *like* and were thus removed from the data set.

(6) *CHI: *like you'd really come (be)cause you work ! (HV5/MT/vicmt5)

Romaine (1991) as well as Meehan (1991) hold that this conjunction variant gave rise to quotative (BE) *like* as in (7), (8), and (9) which were also excluded from further analysis.

(7) *CHI: on the tv you know like I was so scared when the tv was turned off I was like uhh there he is (HV3/ER/mrker3)

(8) *CHI: she's *like* whoa (HV7/MT/sarmt7)

(9) *CHI: and I *like* [=! shriek of fear] (HV3/ER/ether3)

In line with previous research, all the functions of *<like>* listed above were excluded from further analysis. Instances of *<like>* which are considered discourse marker uses are discussed below.
L1-acquisition of the discourse marker *like*

3.3.2 Discourse marker uses

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)). While these criteria were well 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 adverbia (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.

(10)  
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).

(11)  *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
L1-acquisition of the discourse marker like

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 (likecmsm, 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 DM like in speech of the child) or as 1 (DM like was present in the speech of the child). If a child had used DM like in a certain speech situation, it was coded as a DM like user for this situation type at all higher ages. For example, if a child had used DM like during meal time at age 5, it was coded as a DM like user for all subsequent conversations during meal time, i.e. at ages 6 to 11. 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
L1-acquisition of the discourse marker like 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 10.6.

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, given the low overall number of words in some of the situation types and similarities in setting and thus the use of DM like (cf. Figures 1, 2, 3, and 4), the original seven levels were collapse into four levels: experimental task (et), elicited report, reading and letter writing (errlw), meal time (mt), and toy play and book reading (tpbr). The coding of a situation type depended on the information provided in the headers of the HSLLD files. In order to support this re-coding, a cluster analysis was performed on the mean frequencies of DM like by situation type (cf. Figure 10 in the Appendix). The resulting factor has four levels (et, errlw, mt, tpbr,) and is a categorical variable.
3.4.5 Mother's use of the discourse marker like (motlikeptw, independent variable)

The mother's use of DM like represents the overall relative frequency of DM like per 1,000 words based on the total linguistic output 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.

<table>
<thead>
<tr>
<th></th>
<th>LIKEx</th>
<th>Words</th>
<th>Rel. Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(N)</td>
<td>(N)</td>
<td>(ptw)</td>
</tr>
<tr>
<td>Children</td>
<td>307</td>
<td>388,697</td>
<td>0.790</td>
</tr>
<tr>
<td>Mothers</td>
<td>296</td>
<td>753,939</td>
<td>0.393</td>
</tr>
</tbody>
</table>

Figure 3: Relative frequency of DM like by situation type in the HSLLD

Figure 4: Relative frequency of DM like by situation type in the recoded data set
L1-acquisition of the discourse marker like

<table>
<thead>
<tr>
<th>Others</th>
<th>266</th>
<th>234,024</th>
<th>1.137</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>869</td>
<td>1,376,660</td>
<td>2.320</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Age</th>
<th>Sex</th>
<th>Speakers (N)</th>
<th>DM like (N)</th>
<th>Words (N)</th>
<th>Mean DM like (ptw)</th>
<th>Percent (DM like user)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>female</td>
<td>22</td>
<td>0</td>
<td>24,148</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>3</td>
<td>male</td>
<td>29</td>
<td>6</td>
<td>28,998</td>
<td>0.231</td>
<td>12.12</td>
</tr>
<tr>
<td>4</td>
<td>female</td>
<td>41</td>
<td>13</td>
<td>53,083</td>
<td>0.116</td>
<td>19.61</td>
</tr>
<tr>
<td>4</td>
<td>male</td>
<td>32</td>
<td>10</td>
<td>43,229</td>
<td>0.18</td>
<td>21.95</td>
</tr>
<tr>
<td>5</td>
<td>female</td>
<td>35</td>
<td>17</td>
<td>58,336</td>
<td>0.334</td>
<td>31.37</td>
</tr>
<tr>
<td>5</td>
<td>male</td>
<td>31</td>
<td>27</td>
<td>47,743</td>
<td>0.682</td>
<td>35.42</td>
</tr>
<tr>
<td>6</td>
<td>female</td>
<td>6</td>
<td>4</td>
<td>4,761</td>
<td>0.382</td>
<td>33.33</td>
</tr>
<tr>
<td>6</td>
<td>male</td>
<td>5</td>
<td>1</td>
<td>3,558</td>
<td>0.129</td>
<td>33.33</td>
</tr>
<tr>
<td>7</td>
<td>female</td>
<td>34</td>
<td>26</td>
<td>29,534</td>
<td>0.523</td>
<td>38.46</td>
</tr>
<tr>
<td>7</td>
<td>male</td>
<td>32</td>
<td>33</td>
<td>25,749</td>
<td>0.447</td>
<td>31.11</td>
</tr>
<tr>
<td>8</td>
<td>female</td>
<td>2</td>
<td>0</td>
<td>3,833</td>
<td>0.000</td>
<td>33.33</td>
</tr>
<tr>
<td>8</td>
<td>male</td>
<td>1</td>
<td>6</td>
<td>1,279</td>
<td>4.165</td>
<td>100.00</td>
</tr>
<tr>
<td>9</td>
<td>female</td>
<td>20</td>
<td>21</td>
<td>20,206</td>
<td>0.884</td>
<td>46.67</td>
</tr>
<tr>
<td>9</td>
<td>male</td>
<td>19</td>
<td>80</td>
<td>24,977</td>
<td>2.188</td>
<td>51.61</td>
</tr>
<tr>
<td>10</td>
<td>female</td>
<td>9</td>
<td>20</td>
<td>6,681</td>
<td>2.32</td>
<td>33.33</td>
</tr>
<tr>
<td>10</td>
<td>male</td>
<td>9</td>
<td>43</td>
<td>12,582</td>
<td>2.009</td>
<td>61.54</td>
</tr>
<tr>
<td>Total</td>
<td>85</td>
<td>307</td>
<td>388,697</td>
<td>0.491</td>
<td>63.53</td>
<td></td>
</tr>
</tbody>
</table>

Table 2 shows that the final data set consists of 85 speakers, 307 instances of DM like, and 388,697 words. Table 2 shows that the speakers are distributed rather unevenly across ages groups as there are only 3 speakers aged 8. The largest sub-cohort with 73 speakers
L1-acquisition of the discourse marker like consists of children being recorded at age 4. The highest number of instances of DM like within the sub-cohorts is 80 among male children recorded at age 9. Overall, the mean of DM like per 1,000 words in the data amounts to 4.165 which is caused, however, by a single male child using 6 instances of DM like with a conversation encompassing 1,279 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 (combinations of age groups and situation types which did not contain speakers were removed)

<table>
<thead>
<tr>
<th>SitType</th>
<th>Age</th>
<th>Speakers (N)</th>
<th>DM like (N)</th>
<th>Words</th>
<th>Mean DM like (ptw)</th>
<th>Percent (DM like user)</th>
</tr>
</thead>
<tbody>
<tr>
<td>errlw</td>
<td>3-4</td>
<td>75</td>
<td>3</td>
<td>10,818</td>
<td>0.206</td>
<td>3.95</td>
</tr>
<tr>
<td>mt</td>
<td>3-4</td>
<td>68</td>
<td>24</td>
<td>54,078</td>
<td>0.388</td>
<td>21.92</td>
</tr>
<tr>
<td>tpbr</td>
<td>3-4</td>
<td>81</td>
<td>2</td>
<td>83,955</td>
<td>0.015</td>
<td>2.47</td>
</tr>
<tr>
<td>errlw</td>
<td>5-6</td>
<td>67</td>
<td>17</td>
<td>11,809</td>
<td>1.058</td>
<td>14.93</td>
</tr>
<tr>
<td>et</td>
<td>5-6</td>
<td>66</td>
<td>2</td>
<td>17,223</td>
<td>0.112</td>
<td>3.03</td>
</tr>
<tr>
<td>mt</td>
<td>5-6</td>
<td>53</td>
<td>18</td>
<td>35,671</td>
<td>0.506</td>
<td>40.74</td>
</tr>
<tr>
<td>tpbr</td>
<td>5-6</td>
<td>69</td>
<td>12</td>
<td>49,695</td>
<td>0.18</td>
<td>13.04</td>
</tr>
<tr>
<td>errlw</td>
<td>7-8</td>
<td>62</td>
<td>4</td>
<td>11,311</td>
<td>0.2</td>
<td>17.46</td>
</tr>
<tr>
<td>et</td>
<td>7-8</td>
<td>5</td>
<td>5</td>
<td>2,901</td>
<td>1.045</td>
<td>20.00</td>
</tr>
<tr>
<td>mt</td>
<td>7-8</td>
<td>42</td>
<td>52</td>
<td>27,945</td>
<td>1.372</td>
<td>57.14</td>
</tr>
<tr>
<td>tpbr</td>
<td>7-8</td>
<td>62</td>
<td>4</td>
<td>18,238</td>
<td>0.2</td>
<td>17.74</td>
</tr>
<tr>
<td>errlw</td>
<td>9-10</td>
<td>48</td>
<td>19</td>
<td>14,279</td>
<td>0.735</td>
<td>29.17</td>
</tr>
<tr>
<td>et</td>
<td>9-10</td>
<td>50</td>
<td>55</td>
<td>27,449</td>
<td>1.775</td>
<td>44.00</td>
</tr>
<tr>
<td>mt</td>
<td>9-10</td>
<td>39</td>
<td>90</td>
<td>22,718</td>
<td>2.86</td>
<td>84.62</td>
</tr>
<tr>
<td>Total</td>
<td>85</td>
<td>307</td>
<td>388,697</td>
<td>0.491</td>
<td>63.53</td>
<td></td>
</tr>
</tbody>
</table>

Table 3 indicates that the highest frequencies of DM like occur during meal time conversations and that the percentages of children that use DM like increase with age. It also
L1-acquisition of the discourse marker *like* 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.

![Figure 5](image1.png) ![Figure 6](image2.png)

**Figure 5: Percentage of children that have used DM *like* by age**

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 speakers. Figure 6 indicates that male speakers tend to use DM *like* slightly more than female speakers do.
L1-acquisition of the discourse marker *like*

Figure 7 shows that the highest percentages of children using DM *like* occur during meal time conversations and that there appears to be a near-linear, monotonic increase in these percentages. It is remarkable, that between the ages 9 and 10, 84.6 percent of children in the data have at least used DM *like* once during a meal time conversation. In contrast, only 29.2 percent of children aged 9 and 10 in the data have at least used DM *like* once during elicited reports, reading, and letter writing. The percentages of DM *like* users during the experimental task are in-between, i.e. the percentages are higher than during reading and letter writing but substantially lower compared to meal time conversations. It is also remarkable that over a fifth of the children aged 3 and 4 have used DM *like* during meal time conversations.

**Figure 7: Percentage of children that have used DM *like* by age and situation type**

**Figure 8: DM *like* by age and situation type**
L1-acquisition of the discourse marker like

Figure 9 suggests an interaction between the use of DM like by children, their age, and the use of DM like by mothers. During the early stages of acquisition, there appears to be a very strong impact of the mother's use of DM like on the likelihood that a child uses this marker as the probability of a child to use DM like is substantially higher if the mother uses DM like (indicated by the peak of the blue dotted line among children aged 4). However, at age 6, this trend is reversed as the likelihood of a child to use DM like is lower among children whose mother uses DM like frequently (indicated by the red line which shows a higher frequency of DM like users among children whose mother does not use DM like at age 6). During later stages of acquisition, there appears to be no difference between children whose use DM like and children who do not use it with respect to mother's frequency of DM like (indicated by the similarity of the lines from age 7 onward).

Figure 10 indicates an interaction between situation type and the gender of children on the one hand and their likelihood to be a DM like user: while male children are more likely to use DM like during elicited reports, reading, and letter writing as well as during the experimental task, female children are more prone to use DM like during book reading and toy play as well as meal time conversations.
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. 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 5-year old male recorded during toy play.

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.

The step-wise step-up model fitting procedure arrived at a highly significant final minimal adequate model (cf. Tables 6 and 7 in the Appendix). The analysis included only higher-level interactions which did not cause cases of complete separation or lead to failures for the model to converge.

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

The final minimal adequate mixed-effects binomial logistic regression model performs significantly better than a base-line model and reports the age of children, the situation type, and an interaction between situation type and the gender of children as significant
L1-acquisition of the discourse marker like predictors. The model did neither confirm that the gender of children affects the probability of a speaker having used DM like\textsuperscript{13} nor did it confirm a significant impact of the mother's use of DM like on the probability of a speaker having used this particle or on the age with which children become DM like users (cf. Tables 6). In addition, the model did not report any significant interactions except for the interaction between situation type and the gender of children. Including the random effect (random intercepts) was justified as it significantly improved model fit.\textsuperscript{14}

The situation type meal time has the strongest impact on the usage of DM like: during meal time the probability of a child being a DM like user is 36.03 higher than during elicited reports, reading, and letter writing (the baseline or reference category). During book reading and play time, children are also more likely to use DM like compared to the reference category as the probability of a speaker being a DM like user decrease by a factor of 5.33. The final minimal adequate model cannot confirm a significant difference between the experimental task and the reference category (cf. Table 4). Age also substantially affects the likelihood of DM like usage: the probability of speaker becoming a DM like user increases by a factor of 1.84 per year.

The interaction between situation type and the gender of children is relatively easy to interpret. Figure 10 can guide the interpretation: male children are more likely to use DM like during elicited reports, reading, and letter writing as well as during the experimental task, while female children are more prone to use DM like during book reading and toy play as well as meal time conversations.

5 Discussion

The current analysis of the use of DM like during early stages of L1-acquisition in New England English has shown that the usage of DM like is present in the speech of 3- and 4-year
old children and that a third of all children have used it at age 5 (cf. Figure 5). The present study expands previous studies in that it uses data from 85 children which were recorded repeatedly between the ages of 3 and a half and 10 years of age during different situation types. Thus, the data used in this study allowed to investigate the acquisition of an innovative feature of English during early stages of L1-acquisition.

The results of the mixed-effects model show that DM like is used particularly often during family meal time conversations, i.e. a situation type during which investigators were not present. This strongly suggest that children are not only aware of usage constraints at a very young age but that they are also 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 confirms 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, 84.6 percent of children have used DM like during informal meal time conversations when they are 9 or 10 years of age. Furthermore, the probability of a child having used DM like during meal time conversations compared to other situation types 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
L1-acquisition of the discourse marker *like*

children which suggests that their awareness of register variation and thus stylistic constraints on variation 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 (cf. Miller & Weinert, 1995) could be that while DM *like* is acquired very early on, Figure 6 indicates that the frequency of use increases substantially at age 8 or 9. 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 1.84 per year. The percentages of DM *like* users among children appear to increase near-linearly with age (cf. Figure 5). It needs to be born in mind, however, that the percentages relate to whether a child has used DM *like* in a given situation type before as it will be considered a DM *like* user in all subsequent conversations in that situation type in this study. Nonetheless, the results confirm that a substantial percentage of children as young as 3 or 4 years of age have acquired DM *like*. This finding contradicts previous research according to which DM *like* is acquired after an age of 10 (cf. Miller & Weinert, 1995). This finding does however support that research which found that the age between 3 and 4 years is a crucial period 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).
L1-acquisition of the discourse marker like

Furthermore, the present analysis thus casts a shadow of doubt on previous claims according to which DM like is acquired rather late, i.e. at approximately age 10 (Miller & Weinert, 1995). The study does show that DM like is present in the speech of children younger than 4 years of age and that one third of children have acquired it at age 5. This finding challenges the claim that the late acquisition of DM like contradicts a psycholinguistic model of discourse production sketched in Ochs (1979). Rather the data support 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). The findings of the present analysis thus have implications for theories of language acquisition and its relation to discourse production. Interestingly, the results that have been presented are fully in line with construction grammar approaches to language acquisition (cf. e.g. Hilpert, 2014:155-178) which 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, 2009).

Neither the gender of children nor the use of DM like by their mothers correlate significantly with the child's likelihood of becoming a user of DM like. 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, for very young children aged between 3 and 4 years of age, the use of DM like by mothers has a substantial impact on the likelihood of these 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, but the impact of the mothers' frequency of DM like use decreases as children grow older. While the graphical display is
L1-acquisition of the discourse marker like

intriguing as both variationist theories on language transmission (Labov, 1994) and models of language acquisition (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 possible explanation for the lack of a significant interaction between the mothers’ use of DM like and the age at which children acquire DM like relates to the coding of the frequency of DM like use by the mother which was coded as the overall relative frequency of DM like per 1,000 words in the linguistic output of the mother and thus did not reflect situation type specific use of DM like. Arguably, using a situation type specific coding would have been preferable. Unfortunately, the variability in the mothers' output was so substantial in term of quantity that coding the mothers' use of DM like accordingly resulted in unreliable regression models. Whatever the case may be, the visualization in Figure 9 can only lend tentative support to studies which 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. 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 that remains unsolved is 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 could not be addressed here and it would in any case
L1-acquisition of the discourse marker like 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. Further analysis on a wider range of variables is needed to improve our understanding of this shift.

6 Conclusion

This study set out to investigate at what age DM like is acquired and which factors affect the likelihood of children to become DM like users. It has been shown that contrary to Miller & Weinert’s (1995) claim that DM like is acquired “apparently after age 10”, a large proportion of children has acquired DM like substantially earlier (it is present in the speech of 3 and 4 year-olds and around age 5 approximately a third of children in the HSLLD data have made use of DM like). From a theoretical perspective these findings are intriguing as they lend support to previous findings which show that the age of 3 to 4 is crucial with respect to acquiring variation (Roberts & Labov, 1995) and they may be taken to support usage-based models of language acquisition (Tomasello, 2009) - although this framework is fully appropriate given the results as it must be born in mind that the linguistic output of the mother did not significantly correlate with the probability of child to become a DM like user. However, if the latter approach is considered, it can be argued that this study lends credibility to the hypothesis that features acquired early on represent deeply entrenched constructions which speakers heavily rely on during spontaneous speech production as such deeply entrenched construction require less processing to be activated. Nonetheless, the results of the present study support the view - at least with respect to DM like - that highly frequent
L1-acquisition of the discourse marker *like* discourse features should be expected to be acquired during early stages of language acquisition.
L1-acquisition of the discourse marker like

7 Endnotes

---

1 Hereafter DM like.

2 Whenever the sequence like is included in pointy brackets, i.e. <like>, then this refers to the orthographic sequence l-i-k-e irrespective of its function.

3 An overview of the basic characteristics of the data sets used in these studies in provided in Table 5 in the Appendix.

4 According to the respective manual, the HSLLD corpus is said to feature 100 children and covers an age range of 2 to 6 year-olds. However, a closer inspection showed that the HSLLD corpus features only 85 children covering an age range of 3 to 12 year-olds.

5 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).

6 For graphical representations and summaries of the data set, this numeric variable was re-coded as an order factor, i.e. a categorical variable.

7 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 was fitted 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 as in the case of the model reported here.

8 For a more extensive overview of the results cf. Table 4.

9 Model Likelihood Ratio Test: L.R. $\chi^2$: 360.26, DF: 9, Significance: p-value <.001***.

10 $\chi^2$: 132.91, DF: 1, Significance: p-value <.001***.

11 $\chi^2$: 125.78, DF: 2, Significance: p-value <.001***.

12 $\chi^2$: 35.308, DF: 4, Significance: p-value = 0.001***.
While the summary of the final minimal adequate model reports gender as being a significant predictor (z-value: 3.38, p-value <.001***), the more accurate comparison of a model with and a model without gender as a predictor cannot confirm this result ($\chi^2$: 0.67, DF: 1, p-value = 0.41233; cf. also Table 6).

Model Likelihood Ratio Test: L.R. $\chi^2$:66.26, $\chi^2$ DF: 1, Significance: p-value<.001***.
L1-acquisition of the discourse marker like

8 References


L1-acquisition of the discourse marker like


L1-acquisition of the discourse marker like


L1-acquisition of the discourse marker like


L1-acquisition of the discourse marker like


L1-acquisition of the discourse marker like

9 Appendix

Table 4: Results of the final minimal adequate mixed-effects binominal logistic regression model

<table>
<thead>
<tr>
<th>Group(s)</th>
<th>Variance</th>
<th>Std. Dev.</th>
<th>L.R. X2</th>
<th>DF</th>
<th>Pr</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Effect(s)</td>
<td>childid</td>
<td>2.08</td>
<td>1.44</td>
<td>66.26</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

| Fixed Effect(s) | Estimate | VIF | OddsRatio | CI(2.5%) | CI(97.5%) | Std. Error | z     | Pr(>|z|) | Significance |
|------------------|----------|-----|-----------|----------|-----------|------------|-------|---------|--------------|
| (Intercept)      | -7.76    | 0   | 0         | 0        | 0         | 0.7        | -11.11| 0       | p < .001*** |
| ageedit          | 0.61     | 1.35| 1.84      | 1.62     | 2.07      | 0.06       | 9.72  | 0       | p < .001*** |
| stypet           | 0.14     | 3.65| 1.15      | 0.37     | 3.6       | 0.58       | 0.24  | 0.8097  | n.s.         |
| stypmt           | 3.58     | 4.52| 36.03     | 14.52    | 89.39     | 0.46       | 7.73  | 0       | p < .001*** |
| styptpbr         | 1.67     | 3.51| 5.33      | 2.21     | 12.85     | 0.45       | 3.73  | 0       | p < .001*** |
| gendermale       | 1.91     | 2.03| 6.73      | 2.23     | 20.34     | 0.56       | 3.38  | 0       | p < .001*** |
| stypet:gendermale| -0.37    | 3.42| 0.69      | 0.17     | 2.85      | 0.72       | -0.51 | 0.6084  | n.s.         |
| stypmt:gendermale| -2.02    | 3.71| 0.13      | 0.04     | 0.4       | 0.56       | -3.6  | 0       | p < .001*** |
| styptpbr:gendermale| -3.12    | 2.37| 0.04      | 0.01     | 0.15      | 0.63       | -4.99 | 0       | p < .001*** |

Model statistics

<table>
<thead>
<tr>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Groups</td>
</tr>
<tr>
<td>Number of cases in model</td>
</tr>
</tbody>
</table>
L1-acquisition of the discourse marker *like*

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed misses</td>
<td>928</td>
</tr>
<tr>
<td>Observed successes</td>
<td>217</td>
</tr>
<tr>
<td>Residual deviance</td>
<td>751.6</td>
</tr>
<tr>
<td>R2 (Nagelkerke)</td>
<td>0.434</td>
</tr>
<tr>
<td>R2 (Hosmer &amp; Lemeshow)</td>
<td>0.324</td>
</tr>
<tr>
<td>R2 (Cox &amp; Snell)</td>
<td>0.27</td>
</tr>
<tr>
<td>C</td>
<td>0.928</td>
</tr>
<tr>
<td>Somers' Dxy</td>
<td>0.856</td>
</tr>
<tr>
<td>AIC</td>
<td>771.6</td>
</tr>
<tr>
<td>BIC</td>
<td>822.03</td>
</tr>
<tr>
<td>Prediction accuracy</td>
<td>87.34%</td>
</tr>
</tbody>
</table>

*Model Likelihood Ratio Test*

<table>
<thead>
<tr>
<th>L.R. X2:</th>
<th>DF: 9</th>
<th>p-value: 0</th>
<th>sig: p &lt; .001***</th>
</tr>
</thead>
</table>
L1-acquisition of the discourse marker *like*

*Figure 11: Cluster analysis of DM *like* usage by situation type*
**L1-acquisition of the discourse marker like**

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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Region</td>
<td>Glasgow</td>
<td>London</td>
<td>Toronto</td>
<td>London</td>
<td>Northern Ireland</td>
</tr>
<tr>
<td>Words (N)</td>
<td>NA</td>
<td>106,798</td>
<td>NA</td>
<td>70,134</td>
<td>50,915</td>
</tr>
<tr>
<td>Speakers (N)</td>
<td>66</td>
<td>52</td>
<td>52</td>
<td>5</td>
<td>21</td>
</tr>
<tr>
<td>Gender</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>male</td>
<td>female</td>
</tr>
<tr>
<td>Age</td>
<td>8</td>
<td>10</td>
<td>13</td>
<td>10-13</td>
<td>10-12</td>
</tr>
<tr>
<td>Absolute Freq. (N)</td>
<td>2</td>
<td>4</td>
<td>23</td>
<td>244</td>
<td>NA</td>
</tr>
<tr>
<td>Relative Freq. (ptw)</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>2.53</td>
<td>NA</td>
</tr>
</tbody>
</table>
L1-acquisition of the discourse marker *like*

Table 6: Results of the model fitting process (step-wise step-up): mixed-effects binominal logistic regression model

<table>
<thead>
<tr>
<th>Model</th>
<th>Term Added</th>
<th>Compared to...</th>
<th>DF</th>
<th>AIC</th>
<th>BIC</th>
<th>LogLikelihood</th>
<th>Residual Deviance</th>
<th>( \chi^2 )</th>
<th>( \chi^2 )-DF</th>
<th>p-value</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1.glmer</td>
<td>ageedit</td>
<td>m0.glmer</td>
<td>3</td>
<td>918.68</td>
<td>933.81</td>
<td>-456.34</td>
<td>912.68</td>
<td>132.91</td>
<td>1</td>
<td>0</td>
<td>p &lt; .001***</td>
</tr>
<tr>
<td>m2.glmer</td>
<td>gender</td>
<td>m1.glmer</td>
<td>4</td>
<td>920.01</td>
<td>940.18</td>
<td>-456</td>
<td>912.01</td>
<td>0.67</td>
<td>1</td>
<td>0.41233</td>
<td>n.s.</td>
</tr>
<tr>
<td>m3.glmer</td>
<td>motlikeptw</td>
<td>m1.glmer</td>
<td>4</td>
<td>919.33</td>
<td>939.51</td>
<td>-455.67</td>
<td>911.33</td>
<td>1.35</td>
<td>1</td>
<td>0.24565</td>
<td>n.s.</td>
</tr>
<tr>
<td>m4.glmer</td>
<td>styp</td>
<td>m1.glmer</td>
<td>6</td>
<td>798.9</td>
<td>829.16</td>
<td>-393.45</td>
<td>786.9</td>
<td>125.78</td>
<td>3</td>
<td>0</td>
<td>p &lt; .001***</td>
</tr>
<tr>
<td>m5.glmer</td>
<td>ageedit:gender</td>
<td>m4.glmer</td>
<td>7</td>
<td>798.51</td>
<td>833.81</td>
<td>-392.26</td>
<td>784.51</td>
<td>2.39</td>
<td>1</td>
<td>0.12185</td>
<td>n.s.</td>
</tr>
<tr>
<td>m6.glmer</td>
<td>ageedit:motlikeptw</td>
<td>m4.glmer</td>
<td>7</td>
<td>800.75</td>
<td>836.05</td>
<td>-393.37</td>
<td>786.75</td>
<td>0.16</td>
<td>1</td>
<td>0.68986</td>
<td>n.s.</td>
</tr>
<tr>
<td>m7.glmer</td>
<td>gender+motlikeptw</td>
<td>m4.glmer</td>
<td>8</td>
<td>801.72</td>
<td>842.07</td>
<td>-392.86</td>
<td>785.72</td>
<td>1.18</td>
<td>2</td>
<td>0.55403</td>
<td>n.s.</td>
</tr>
<tr>
<td>m8.glmer</td>
<td>gender+gender:motlikeptw</td>
<td>m4.glmer</td>
<td>9</td>
<td>803.65</td>
<td>849.03</td>
<td>-392.82</td>
<td>785.65</td>
<td>1.26</td>
<td>3</td>
<td>0.73911</td>
<td>n.s.</td>
</tr>
<tr>
<td>m9.glmer</td>
<td>gender</td>
<td>m4.glmer</td>
<td>7</td>
<td>800.31</td>
<td>835.61</td>
<td>-393.16</td>
<td>786.31</td>
<td>0.59</td>
<td>1</td>
<td>0.44166</td>
<td>n.s.</td>
</tr>
<tr>
<td>m10.glmer</td>
<td>gender+gender:styp</td>
<td>m4.glmer</td>
<td>10</td>
<td>771.6</td>
<td>822.03</td>
<td>-375.8</td>
<td>751.6</td>
<td>35.31</td>
<td>4</td>
<td>0</td>
<td>p &lt; .001***</td>
</tr>
</tbody>
</table>