Ruhr-Universität Bochum

Linguistic Data Science

Usage of German articles in child language acquisition

Submitted by: Sadhana Muthukumar, Master's student in Linguistic Data Science

Supervisor: Dr. Claudia Roch, M.A.

Statutory declaration:

I, *Sadhana Muthukumar*, confirm that the work for the documentation with the title: "Usage of German articles in child language acquisition" was solely undertaken by myself and that no help was provided from other sources as those allowed. All sections of the paper that use quotes or describe an argument or concept developed by another author have been referenced, including all secondary literature used, to show that this material has been adopted to support my work.

Place / Date: Bochum, Germany / 15.05.2024

Signature: *M.Sadhana*

Abstract

In general, children omit or use very less articles during their early language development. This is applicable for German language as well. Moreover, the acquisition of German articles is a challenging task during childhood. It is due to the complex nature of German articles. Some of the main reasons for their complexity are gender assignment and high-level syncretism. In gender assignment, one has to learn the correct gender of a noun and select article in accordance with noun's gender. Also, there are no clear rules about how a noun's gender can be identified in German. Second, the syncretism is again making acquisition process more difficult i.e., the same form of article is used for more than one case, e.g., 'der' is the form for both masculine nominative and feminine dative. These might lead children learning German to face challenges in acquiring and using articles. These challenges have led a way to the existence of this project. The main aim of this project is to investigate whether it is really challenging for children to acquire German articles or not by analyzing the overall quantity of articles' usage during their entire childhood. In addition, it also analyzes which articles are used often/less in which childhood stage and shows how articles' usage are getting transitioned from one stage to another. These analyzes were performed using multiple correspondence analysis (MCA) technique through R programming.

Introduction

In German language, articles play an important role, since they represent the number, gender, and case of a noun[1]. The basic forms of article in German can be divided into two major groups i.e., definite and indefinite article. Der (masculine), die (feminine, plural), and das (neuter) are definite. Ein (masculine, neuter) and eine (feminine) are indefinite. These forms change according to number, gender, and case of a noun. Fig 1.1. shows the entire forms of German article[2].

	indefinite			definite			
	singular			singular			plural
	masculine	feminine	neuter	masculine	feminine	neuter	
nominative	ein	eine	ein	der	die	das	die
accusative	einen	eine	ein	den	die	das	die
dative	einem	einer	einem	dem	der	dem	den
genetive	eines	einer	eines	des	der	des	der

Fig1.1. Forms of German article

[1] In addition to articles, adjectives and possessive nouns are also representing noun's no., gender and case. But, articles are being the basic marker for nouns.

[2] For this project, only nominative, accusative, and dative cases were considered. But, genitive case and plural forms were not considered in order to make the project clear and simple. Also, children will not use genitive case mostly during their early stages of language development.

The omission of articles in child language development is a well-known phenomenon and there are several literature works related to it (Clahsen (1984), Bittner (1997) and so on). This project investigates this phenomenon by researching the usage of German articles by children during their language acquisition in childhood via addressing three main research questions, and they are as follows.

1. *At what age, children start to use articles in general? From when they start to use accusative and dative?*
2. *Which articles' usage is high during childhood: Is it definite or indefinite?*
3. *What is the overall usage of articles among children, is it more or less (by comparing with adults' usage)?*

In this project, only the ages from 1 to 8 years were considered as childhood. In order to analyze the research questions, these ages were divided into five stages as shown in fig 1.2.

stage 1	1 to 2 years
stage2	2 to 3 years
stage3	3 to 4 years
stage4	4 to 6 years
stage5	6 to 8 years

Fig 1.2. Stages and their corresponding age interval

This documentation consists of three major sections. The first section talks about the previous literature works related to this project, second section is about the data used in this project, and in third section, these 3 research questions are analyzed with the help of MCA technique using R programming.

1. Background literature

Some of the main literatures that are related to research questions of this project has been mentioned here. It is important to find out if the findings of this project are in accordance with below statements.

According to Collings (1990), children started using articles from the age of 2 and 1/2 years. At this stage, they mostly use nominative case. They will start using accusative and dative cases from age of 3 and ½ years.

Mills (1986) stated that children produced indefinite article earlier than definite one and the usage of indefinite feminine article 'eine' is very high until age of 3 years.

According to Kupisch et.al. (2009), when the Mean Length of Utterence (MLU) i.e., no. of words in the utterance is above 3, children (including German speaking kids) use articles in more than 80 percent of all contexts. They have also mentioned that the usage of definite article is more compared to indefinite article.

As per Bittner (1997), German-speaking children tend to omit articles in their very early stages of language development especially until 3 years old.

2. CHILDES corpus

This project makes use of data from CHILDES corpus for its work. Child Language Data Exchange System, otherwise known as CHILDES (MacWhinney & Snow, 1985; MacWhinney, 2000, 2014) is a corpus which contains thousands of children language transcripts across 20+ language (including German). In CHILDES, the majority of transcripts are based on interactions and conversations. Childes-db (Sanchez et.al., 2019) is a database-formatted mirror of CHILDES in order to improve the accessibility of this corpus. With the help of 'childesr' package in R, childes-db can be accessed and manipulated.

For this project, a new dataset has been created from childes-db by filtering only German data. 'Childes.Ger.article.df' is the main dataset for this project[3]. It contains data from Caroline, Grimm, Leo, Miller, Rigol, Stuttgart, Szagun, TAKI and Wagner corpora. Fig 2.1 shows sample data of this dataset. It consists of 6,12,683 records (rows) and 7 columns. The columns are 'article', 'speaker_role', 'target_child_name', 'target_child_age', 'corpus_name', 'target_child_age_inyears', and 'stage'. From this main dataset, a sub-dataset has been created for each research question and they will be explained in upcoming section.

article <chr>	speaker_role <chr>	target_child_name <chr>	target_child_age <int>	corpus_name <chr>	target_child_age_inyears <dbl>	stage <chr>
das	Investigator	Andreas	25	Wagner	2.1	stage2
das	Investigator	Andreas	25	Wagner	2.1	stage2
das	Investigator	Andreas	25	Wagner	2.1	stage2
das	Mother	Andreas	25	Wagner	2.1	stage2
das	Investigator	Andreas	25	Wagner	2.1	stage2
das	Investigator	Andreas	25	Wagner	2.1	stage2

Fig 2.1. Main dataset of this project

3. Analyzing research questions using MCA

Correspondence analysis is an exploratory space reduction technique for data analysis. It identifies the patterns of association and disassociation in data. It presents result in the form of two-dimensional plot, which shows the relationship between variables in an intuitive manner. Correspondence analysis can be divided into two major classes. They are simple correspondence analysis (CA), and multiple correspondence analysis (MCA). CA is used to analyze the relationship between two variables, whereas MCA is used for analyzing more than two variables. CA can be implemented only for contingency table i.e., table with cross tabulation showing the frequency distribution of variables. But, MCA can be implemented for raw dataset and it doesn't need to be converted into contingency table.

[3] All the datasets and R markdown files related to this project can be found at https://github.com/velsadhana/EDA-task---CHILDES

Since the dataset of this project is not a contingency table, MCA has been used throughout. In R, there are several packages available for performing MCA. For all the research questions, 'FactoMineR' package (Le et.al., 2008) was used to perform MCA.

3.1. Research question1 (Rq1)

At what age, children start to use articles in general? From when they start to use accusative and dative?

3.1.1. Implementing MCA

'Childes.Ger.article.df1' is the dataset for research question1. It contains 1,72,657 records (rows) and 2 columns such as 'article' and 'stage'. Fig 3.1 shows sample data of this dataset.

article <chr>	stage <chr>
das	stage2
das	stage2
das	stage2
das	stage2
das	stage2
das	stage2

Fig 3.1. Dataset of Rq1

mca.1 <- MCA(childes.Ger.article.df1)

```
Call:
MCA(X = childes.Ger.article.df1)


Eigenvalues
                       Dim.1   Dim.2   Dim.3   Dim.4   Dim.5   Dim.6   Dim.7   Dim.8   Dim.9  Dim.10  Dim.11  Dim.12  Dim.13
Variance               0.545   0.534   0.525   0.509   0.500   0.500   0.500   0.500   0.500   0.491   0.475   0.466   0.455
% of var.              8.390   8.223   8.082   7.835   7.692   7.692   7.692   7.692   7.692   7.550   7.302   7.162   6.994
Cumulative % of var.   8.390  16.614  24.696  32.531  40.223  47.915  55.607  63.300  70.992  78.542  85.844  93.006 100.000

Individuals (the 10 first)
          Dim.1    ctr   cos2    Dim.2    ctr   cos2    Dim.3    ctr   cos2
1       | -0.069  0.000  0.003 |  0.361  0.000  0.070 |  0.109  0.000  0.006 |
2       | -0.069  0.000  0.003 |  0.361  0.000  0.070 |  0.109  0.000  0.006 |
3       | -0.069  0.000  0.003 |  0.361  0.000  0.070 |  0.109  0.000  0.006 |
4       | -0.069  0.000  0.003 |  0.361  0.000  0.070 |  0.109  0.000  0.006 |
5       | -0.069  0.000  0.003 |  0.361  0.000  0.070 |  0.109  0.000  0.006 |
6       | -0.069  0.000  0.003 |  0.361  0.000  0.070 |  0.109  0.000  0.006 |
7       | -0.069  0.000  0.003 |  0.361  0.000  0.070 |  0.109  0.000  0.006 |
8       | -0.069  0.000  0.003 |  0.361  0.000  0.070 |  0.109  0.000  0.006 |
9       | -0.069  0.000  0.003 |  0.361  0.000  0.070 |  0.109  0.000  0.006 |
10      | -0.069  0.000  0.003 |  0.361  0.000  0.070 |  0.109  0.000  0.006 |

Categories (the 10 first)
           Dim.1      ctr     cos2   v.test      Dim.2      ctr     cos2   v.test      Dim.3      ctr     cos2   v.test
das     |   0.642   10.366    0.156  164.051 |   0.477    5.835    0.086  121.843 |   0.307    2.469    0.036   78.579 |
dem     |   0.948    1.298    0.014   49.845 |  -2.464    8.950    0.097 -129.548 |   3.655   20.027    0.214  192.124 |
den     |   1.067    5.905    0.068  108.570 |   0.620    2.036    0.023   63.119 |  -0.334    0.602    0.007  -34.032 |
der     |  -0.082    0.097    0.001  -14.782 |  -0.797    9.504    0.121 -144.504 |  -0.576    5.044    0.063 -104.364 |
die     |   0.049    0.051    0.001   11.237 |  -0.387    3.278    0.046  -88.843 |  -0.543    6.555    0.090 -124.549 |
ein     |  -1.320   27.356    0.360 -249.321 |   0.398    2.535    0.033   75.130 |   0.303    1.494    0.019   57.191 |
eine    |  -0.723    2.904    0.034  -76.295 |   0.291    0.479    0.005   30.674 |   0.265    0.404    0.005   27.926 |
einem   |  -1.007    0.218    0.002  -20.287 |  -3.243    2.307    0.025  -65.335 |   7.577   12.816    0.135  152.657 |
einen   |   0.941    1.446    0.016   52.652 |  -1.125    2.110    0.023  -62.966 |   0.207    0.073    0.001   11.612 |
einer   |   0.712    0.358    0.004   26.082 |   4.239   12.966    0.140  155.299 |   0.838    0.515    0.005   30.695 |

Categorical variables (eta2)
          Dim.1 Dim.2 Dim.3
article | 0.545 0.534 0.525 |
stage   | 0.545 0.534 0.525 |
```

Fig 3.2. Summary of MCA for Rq1

Fig 3.2 shows the result of MCA summary. It can be divided into 4 parts such as eigenvalues, individuals, categories, and categorical variables.

Eigenvalues represent the variance of data in each dimension. The variance value for first 2 dimensions is always greater than other dimensions. In this result, the first two dimensions together represent 16% of the total variance of data. In general, variance depends on the size of data. When the data size is large, there are chances for the less variance percentage in dimension. Since our data is large, it is not possible to show the entire variance in first 2 dimensions itself. But, it is still recommended to consider, as the first 2 dimensions provide the maximum variance as much as possible.

Individuals part provides details about each row and it is not necessary for this work. Hence, it will not be discussed.

In categories part, cos2 (cosine square) represents the quality of each category. In the first dimension, 'ein' has the highest cos2 value and in the second dimension, 'einer' has highest value. The column 'ctr' provides the percentage of contribution of each category in every dimension. When a category has high percentage, then it contributes the most to their corresponding dimension. In the first dimension, only 'ein' is having high contribution and in the second dimension, only 'einer' is having high contribution.

The last part 'categorical variables' will not be discussed. It contains only the coordinate values of the variables, which is not necessary for this work.

'Factoextra' is a package which provides an elegant and clear visualization of MCA graph without any overlapping. Fig 3.3 shows the graph of MCA using factoextra package.

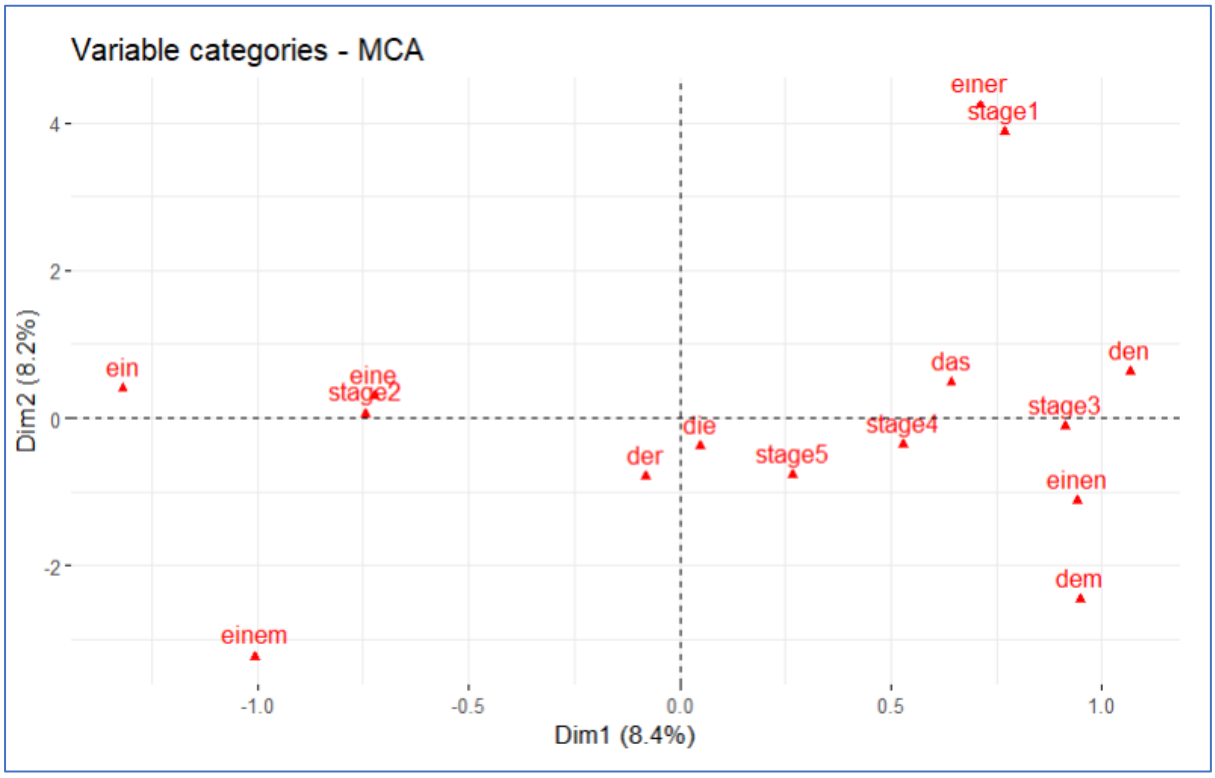

Fig 3.3. MCA graph for Rq1

In fig 3.3, stage1 is separated from stage2. In general, when categories from same variable are separated from each other, they are having different profiles. Similarly, stage3, stage4, and stage5 are separated from both stage1 and stage2. But stage3, stage4, and stage5 are somehow associated and there is not much deviation between them. Since these categories are quite grouped, they have similar profile.

The articles 'ein', 'einem' are far separated from all other categories, meaning they are not in much association with any of the stages. 'Die' and 'der' are close to each other, 'das' and 'den' are slightly close to each other.

Relationship between stage and article: 'einer' is closely associated with stage1, 'eine' is closely associated with stage2. Likewise, 'den', 'das', and 'einen' are grouped with stage3, also 'dem' is slightly close to stage3. 'das' lies in between stage3 and stage4 and this implies that 'das' is used in both stage3 and stage4. However, it is very close to stage4. 'der' and 'die' are closely associated with stage5.

3.1.2. Answering research question 1

At what age, children start to use articles in general?

From above MCA interpretation, it is clear that children are started to use articles at the age of 1 itself. Because, stage1 (1 to 2 years) is associated with an article ('einer'). This answer is in contrast with Collings (1990). According to him, children started using articles at the age of 2 and ½ years.

Moreover, stage2 (2 to 3 years) is closely associated with ‘eine’ and this is in accordance with Mills (1986) as he mentioned 'eine' is very high until age of 3 years.

From when they start to use accusative and dative?

In German, there is a high level of syncretism with article forms i.e., the same form of article is used for more than one case. For example: Definite article 'die' is the form for both feminine nominative and feminine accusative, 'das' is the form for both neuter nominative and neuter accusative. Likewise, indefinite article 'eine' is the form for feminine nominative and feminine accusative, and so on.

The same problem has been mentioned in Collings (1990) as well. *"Das for neuter nominative cannot be distinguished from das neuter accusative, just as die feminine nominative cannot be distinguished from die feminine accusative because there is no phonological difference; der, on the other hand, for masculine nominative can easily be distinguished from its morphological counterpart in the accusative, den. Researchers like Clahsen have argued that the emergence of accusative can only be observed when limited to den, the only case marking being visible. As for neuter and feminine, one cannot say for sure if the child uses nominative or accusative"* (Collings, 1990).

In this project as well, the above facts were taken into consideration. So, the usage of accusative is identified by only observing 'den'/'einen', and dative by 'dem'/'einem', as they are the good indicators for accusative and dative respectively.

Coming to the answer, children will start using accusative and dative from the age of 3 years. Because, 'den' and 'einen' are closely associated with stage3 (3 to 4 years), and 'dem' is also slightly close to stage3. This analysis is somehow matching with Collings (1990), as he pointed out that the usage of accusative and dative begins at age of 3 and ½ years.

Although, they started to use both accusative and dative from 3 years old, their fluency with accusative is higher than dative. For dative, they require more time to get a grip of it and 3 years is too early. In order to test this, some sample data has been taken from Miller corpus (within CHILDES). This data contains conversational transcripts of a child named 'Kerstin' who was 3 years old.

Transcripts with accusative 'den':

Investigator	tun wir den Peter noch ein+bisschen stören
Target_Child	aber den kriege ich wieder
Investigator	du frag erst mal den Peter ob du da was nehmen kannst
Investigator	und was haste mit den Fischen gemacht
Target_Child	ja das könnt man den da haben
Investigator	ich fang den gelben Fisch ja
Investigator	und warum darf ich denn den gelben Fisch nicht fangen
Investigator	ich möcht doch auch den gelben Fisch fangen
Investigator	warum darf ich den nicht fangen
Investigator	warum hast 'n den weggeworfen
Target_Child	den schmeiss auch weg
Target_Child	die fang ich auch immer den da
Target_Child	den da muss auch mitfahren ja
Target_Child	kannst du auch auf den Topf hinsetzen
Investigator	ob ich auch auf den Topf sitzen kann

Transcripts with dative 'dem':

Father	da hm nimm einmal die Flasche aus dem Mund
Investigator	nee da tuste davor stehen vor dem Stuhl
Target_Child	die Maria kann ja dem
Investigator	mit dem Blei kann man sich nicht vollmalen
Target_Child	aber mit dem da kannst du malen
Investigator	sage mal Kerstin seid ihr mit dem Auto weggefahren
Investigator	wer ist denn da alles mitgefahren mit dem Auto du bist doch mit dem Papa und mit der Mama mit dem Auto weggefahren
Investigator	ja
Target_Child	mit dem xxx tragen
Target_Child	und dann muss ich mit dem da brauchen
Investigator	gucke mal da sind Fische draufgemalt auf dem Aquarium siehste
Investigator	ja jetzt guck jetzt hab ich neben dem Kopf
Mother	das gehört zu dem Auto da
Target_Child	aber mit dem da
Investigator	mit dem kannst du schon fahren ne
Investigator	mit dem Auto vom Karussell kannst du schon fahren

From the above conversations, it is clear that the child is having more knowledge with accusative than dative. Because, the child's usage of 'den' is frequent and correct, but the usage of 'dem' is not so often as well as not providing correct meaning of dative.

3.2. Research question2 (Rq2)

Which articles' usage is high during childhood: Is it definite or indefinite?

3.2.1. Implementing MCA

'childes.Ger.article.df2' is the dataset for research question 2. It contains 1,72,657 rows and 2 columns such as 'stage' and 'def' (definite/indefinite). Fig 3.4. shows sample data of this dataset.

stage <chr>	def <chr>
stage2	definite
stage2	definite
stage2	definite
stage2	definite
stage2	definite
stage2	definite

Fig 3.4. Dataset of Rq2

mca.2 <- MCA(childes.Ger.article.df2)

```
Call:
MCA(X = childes.Ger.article.df2)


Eigenvalues
                       Dim.1   Dim.2   Dim.3   Dim.4   Dim.5
Variance               0.537   0.500   0.500   0.500   0.463
% of var.             21.487  20.000  20.000  20.000  18.513
Cumulative % of var.  21.487  41.487  61.487  81.487 100.000

Individuals (the 10 first)
           Dim.1    ctr   cos2    Dim.2    ctr   cos2    Dim.3    ctr   cos2
1          | 0.165  0.000  0.038 | -0.072  0.000  0.007 | -0.220  0.000  0.068 |
2          | 0.165  0.000  0.038 | -0.072  0.000  0.007 | -0.220  0.000  0.068 |
3          | 0.165  0.000  0.038 | -0.072  0.000  0.007 | -0.220  0.000  0.068 |
4          | 0.165  0.000  0.038 | -0.072  0.000  0.007 | -0.220  0.000  0.068 |
5          | 0.165  0.000  0.038 | -0.072  0.000  0.007 | -0.220  0.000  0.068 |
6          | 0.165  0.000  0.038 | -0.072  0.000  0.007 | -0.220  0.000  0.068 |
7          | 0.165  0.000  0.038 | -0.072  0.000  0.007 | -0.220  0.000  0.068 |
8          | 0.165  0.000  0.038 | -0.072  0.000  0.007 | -0.220  0.000  0.068 |
9          | 0.165  0.000  0.038 | -0.072  0.000  0.007 | -0.220  0.000  0.068 |
10         | 0.165  0.000  0.038 | -0.072  0.000  0.007 | -0.220  0.000  0.068 |

Categories
             Dim.1      ctr     cos2   v.test      Dim.2      ctr     cos2   v.test      Dim.3      ctr     cos2   v.test
stage1     |   0.516    0.769   0.009   38.364 |  -0.449    0.627   0.006  -33.422 |  -0.437    0.593   0.006  -32.515
stage2     |   0.676   20.574   0.428  271.706 |  -0.102    0.506   0.010  -41.124 |  -0.312    4.692   0.091 -125.177
stage3     |  -1.021   24.139   0.345 -244.113 |  -0.906   20.410   0.272 -216.564 |  -0.496    6.110   0.081 -118.489
stage4     |  -0.534    4.513   0.058 -100.418 |   2.068   72.639   0.875  388.692 |   0.263    1.174   0.014   49.413
stage5     |   0.028    0.005   0.000    3.135 |  -0.929    5.817   0.062 -103.777 |   3.602   87.431   0.937  402.320
definite   |  -0.434   12.982   0.537 -304.540 |   0.000    0.000   0.000    0.000 |   0.000    0.000   0.000    0.000
indefinite |   1.238   37.018   0.537  304.540 |   0.000    0.000   0.000    0.000 |   0.000    0.000   0.000    0.000

stage1     |
stage2     |
stage3     |
stage4     |
stage5     |
definite   |
indefinite |

Categorical variables (eta2)
           Dim.1 Dim.2 Dim.3
stage    | 0.537 1.000 1.000 |
def      | 0.537 0.000 0.000 |
```

Fig 3.5. Summary of MCA for Rq2

In fig 3.5, the first two dimensions together represent 41% of the total variance of data.

In categories part, 'definite', 'indefinite', and 'stage2' are categories with good quality for first dimension as they have high cos2 value. In second dimension, only 'stage4' has high cos2 value. In first dimension, 'indefinite' is having high contribution as their ctr value is high. In second dimension, only 'stage4' has very high contribution.

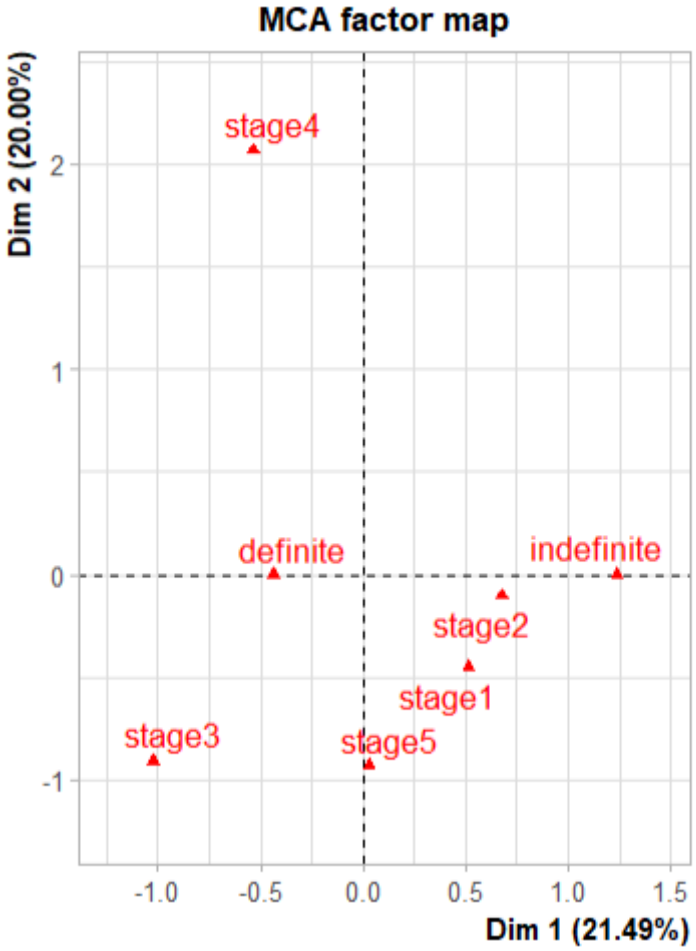

Fig 3.6. MCA graph of Rq2

In fig 3.6, stage4 is far separated from all the other categories. Stage1 and stage2 are closely associated, stage5 is slightly associated with stage1.

The definite and indefinite articles are separated from each other. It means, they are having different profiles.

Relationship between stage and definite/indefinite article: 'Indefinite' is closely associated with stage2 and stage1. But, the definite article is not closely associated with any of the stages.

3.2.2. Answering research question2

Which articles' usage is high during childhood: Is it definite or indefinite?

In general, children are abundantly using definite article during their childhood, as the frequency of definite is very high compared to indefinite (please see fig 3.7). But, during very early stages, they are using indefinite article more than definite one, since indefinite is grouped with stage1 and stage2. In the sense, their usage of indefinite is very high between 1 to 3 years (stage1 & stage2). This is in accordance with the work of Mills (1986), as he stated that children use indefinite earlier than definite.

As per fig 3.6, definite is not at all closely associated with any of the stages. It means, the usage of definite article is not exceptionally high at any particular stage. The definite lies approximately in the center to stage3, stage5, stage1, and stage2, so one can say that definite article is equally distributed among these stages. Due to the far separation of stage4, it is difficult to make predictions out of it.

During early childhood, they use indefinite article quite often until 3 years old. Afterwards, the usage shifted from indefinite to more definite article. Overall, children are using definite articles in a high level during their entire childhood.

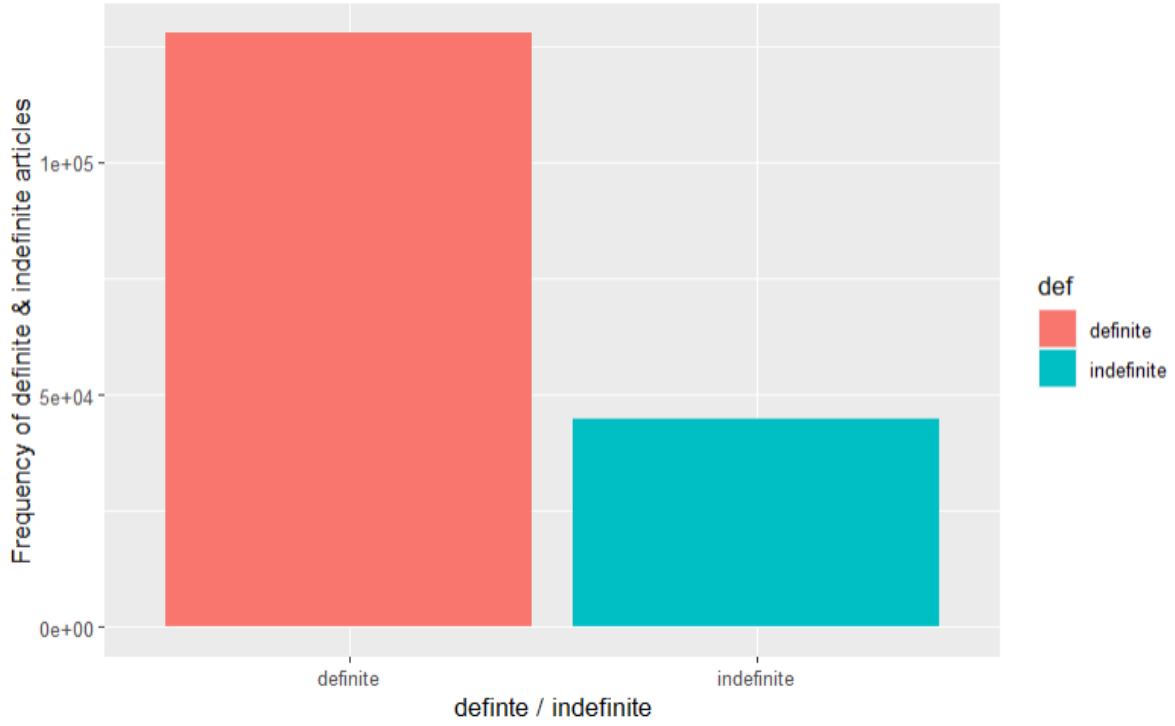

Fig 3.7. Bar chart representing frequency of definite and indefinite articles

3.3. Research question3 (Rq3)

What is the overall usage of articles among children, is it more or less (by comparing it with adults' usage)?

3.3.1. Implementing MCA

'childes.Ger.article.df3' is the dataset for research question 3 and it contains 6,08,304 rows and 2 columns such as ‘article’ and ‘speaker’. Fig 3.8 shows sample data of this dataset.

article <chr>	speaker <chr>
definite	adults
definite	adults
definite	adults
definite	adults
definite	adults
definite	adults

Fig 3.8. Dataset of Rq3

mca.3 <- MCA(childes.Ger.article.df3)

```
Call:
MCA(X = childes.Ger.article.df3)


Eigenvalues
                       Dim.1   Dim.2
Variance               0.528   0.472
% of var.             52.820  47.180
Cumulative % of var.  52.820 100.000

Individuals (the 10 first)
              Dim.1    ctr   cos2    Dim.2    ctr   cos2
1           | -0.582  0.000  0.993 |  0.047  0.000  0.007 |
2           | -0.582  0.000  0.993 |  0.047  0.000  0.007 |
3           | -0.582  0.000  0.993 |  0.047  0.000  0.007 |
4           | -0.582  0.000  0.993 |  0.047  0.000  0.007 |
5           | -0.582  0.000  0.993 |  0.047  0.000  0.007 |
6           | -0.582  0.000  0.993 |  0.047  0.000  0.007 |
7           | -0.582  0.000  0.993 |  0.047  0.000  0.007 |
8           |  0.527  0.000  0.198 | -1.062  0.000  0.802 |
9           | -0.582  0.000  0.993 |  0.047  0.000  0.007 |
10          |  0.527  0.000  0.198 | -1.062  0.000  0.802 |

Categories
               Dim.1      ctr     cos2   v.test      Dim.2      ctr     cos2   v.test
definite    |  -0.389   11.120    0.528 -566.839 |  -0.367   11.120    0.472 -535.721 |
indefinite  |   1.359   38.880    0.528  566.839 |   1.284   38.880    0.472  535.721 |
adults      |  -0.458   14.192    0.528 -566.839 |   0.432   14.192    0.472  535.721 |
children    |   1.154   35.808    0.528  566.839 |  -1.091   35.808    0.472 -535.721 |

Categorical variables (eta2)
             Dim.1 Dim.2
article     | 0.528 0.472 |
speaker     | 0.528 0.472 |
```

Fig 3.9. Summary of MCA for Rq3

In fig 3.9, the first two dimensions together represent 100% of the total variance of data!

In categories part, all the 4 categories i.e., 'definite', 'indefinite', and 'adults' and 'children' are having same cos2 value (0.528) in first dimension. So, they provide the same level of quality to the first dimension. Similarly, in second dimension, again all these categories are having same value (0.472). In the first dimension, 'indefinite' is having high contribution as their ctr value is high compared to others. In second dimension as well, 'indefinite' provides the high contribution.

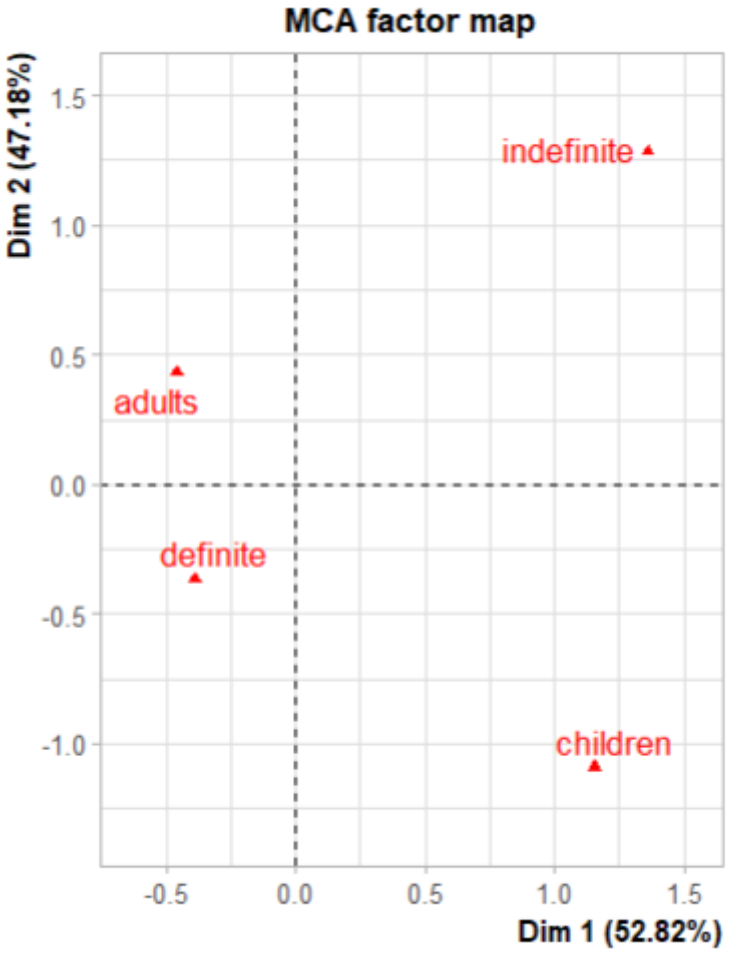

Fig 3.10. MCA graph of Rq3

In the fig 3.10, indefinite is far separated from all the other categories. The categories 'definite' and 'indefinite' are separated from each other, such that they are having different profiles.

The category 'children' is highly separated from all the other categories. Also, children and adults are separated from each other in the opposite quadrants. Hence, they are having different profiles.

Relationship between speaker and article: There are not much association between any of the categories other than ‘adults’ and ‘definite’. These two categories are somehow associated with each other. The separation of ‘indefinite’ shows that it is not much used by both adults and children.

3.3.2. Answering research question3

What is the overall usage of articles among children, is it more or less (by comparing it with adults' usage)?

In general, the usage of articles by children is always very less. But, how one could say the usage is less? One of the best ways is to compare it with adults. The same method has been applied here. In fig 3.10, 'children' is far separated from all the other categories, and it depicts that they are using very less amount of both articles comparing to adults. 'Definite' is slightly close to 'adults' and 'indefinite' is quite far from it, meaning adults are using more definite article than indefinite.

The below bar chart (fig 3.11) also shows that the overall usage of articles is less among children in comparing to adults.

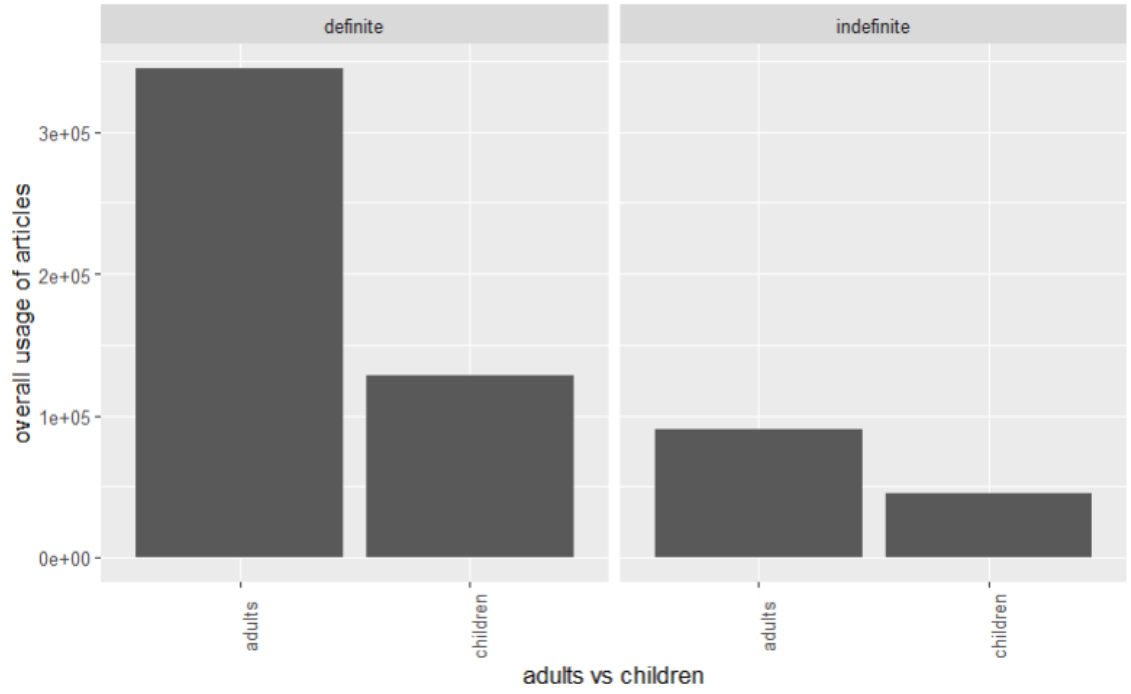

Fig 3.11. Bar chart representing usage of articles by adults vs children

Conclusion

This project highlights some critical findings and they are as follows. Children who learn German are starting to use articles at the age of 1 itself. At the age of 3 years, they begin to use dative and accusative, in the sense they are able to differentiate nominative and accusative/dative. In childes corpus, adults have used articles 4,35,647 times, and children used 1,72,657 times. In other words, adults have used 72% of articles and children have used only 28%. Therefore, it is clear that children are using very less number of articles during their childhood. The most important thing to be noted here is, all these findings are based on average perspective and it cannot be true for all children, i.e., the acquisition of German articles depends on the capability of each and every children. For e.g., one child

might use articles at the age of 1 itself and the other might use articles at the age of 3 or 4 years. Furthermore, these findings are the reflection of CHILDES corpus data. Hence, they may vary with other corpora. In order to get more perfect result, it is always recommended to work with larger data from various corpora as much as possible (although CHILDES corpus is well-known for children language transcripts). The more vast the data is, the more perfect the result will be. Finally, I am concluding my analysis with a statement that, "according to CHILDES corpus, the overall quantity of articles' usage by children (from 1 to 8 years) is very less and this indirectly signifies the complexity of German articles during early language acquisition."

References

Bittner, D. (1997). Entfaltung grammatischer Relationen im NP-Erwerb: Referenz. *Folia Linguistica*, 31(3–4), 255–283.

Clahsen, H. (1984). Der Erwerb von Kasusmarkierungen in der deutschen Kindersprache. *Linguistische Berichte*, 89, 1–31.

Collings, A. (1990). The acquisition of morphology and syntax in German child language.

Fuchs, J., Domahs, U., & Kauschke, C. (2021). Information structure in language acquisition. Production and comprehension of (in) definite articles by German-speaking children. *Journal of Child Language*, 48(1), 55-87.

Glynn, D. (2014). Correspondence Analysis. Exploring data and identifying patterns. In Glynn, D. & Robinson, J. A. (eds.) *Corpus methods for Semantics. Quantitative studies in polysemy and synonymy,* (pp. 443-486). Amsterdam/Philadelphia: John Benjamins.

Husson, F., Lê, S., & Pagès, J. (2011). *Exploratory multivariate analysis by example using R* (Vol. 15). Boca Raton: CRC press.

Kupisch, T., Anderssen, M., Bohnacker, U., & Snape, N. (2009). Article acquisition in English, German, Norwegian and Swedish. *Little words: Their history, phonology, syntax, semantics, pragmatics and acquisition (Proceedings of The Georgetown University Round Table,* GURT 2007), 223-236.

Le, S., & Husson, F. (2008). FactoMineR: An R package for multivariate analysis. *Journal of Statistical Software,* 25, 1–18.

MacWhinney, B. (2000). *The childes project: The database (Vol. 2).* Psychology Press.

MacWhinney, B. (2014). *The childes project: Tools for analyzing talk, volume ii: The database.* Psychology Press.

MacWhinney, B., & Snow, C. (1985). The child language data exchange system. *Journal of Child Language, 12 (2),* 271–295.

Mills, A. E. (1986). *The acquisition of gender. A study of English and German*. Berlin: Springer.

R Core Team. (2017). R: *A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Sanchez, A., Meylan, S. C., Braginsky, M., MacDonald, K. E., Yurovsky, D., & Frank, M. C. (2019). childes-db: A flexible and reproducible interface to the child language data exchange system. *Behavior research methods, 51,* 1928-1941.