



# THE INFLUENCE OF EXPOSURE AND AGE ON IDIOM FAMILIARITY IN DUTCH CHILDREN

Bachelor's Project Thesis

Maaïke Looijenga, s4812328, m.looienga.1@student.rug.nl,

Supervisor: A.R. Thiel

**Abstract:** This study investigates the familiarity and comprehension of idiomatic expressions among Dutch children aged 7 to 11. Utilizing a multiple-choice questionnaire, 44 participants were assessed on their recognition of 26 idioms commonly found in Dutch children's literature. The idioms were selected based on a frequency analysis from a corpus study of Dutch children's books. Results indicate a positive relationship between idiom frequency and familiarity. There is no significant evidence that older children demonstrate a higher recognition rate of idiomatic expressions compared to younger participants. The findings emphasize the role of idiom frequency in language acquisition, suggesting that frequent exposure to idiomatic expressions can enhance familiarity, irrespective of age. This research contributes to a nuanced understanding of language development, highlighting the importance of idiomatic knowledge in linguistic and cognitive growth.

## 1 Introduction

In our daily conversations, we frequently use idioms to convey complex ideas and emotions succinctly and effectively. Idioms are multiword expressions imbued with figurative meanings (Oxford English Dictionary, 2023). They constitute a fundamental aspect of language, enriching communication with nuance and cultural context. An example is 'break a leg'; used when wishing someone good luck, not telling them to literally break one of their legs.

Idioms come in different forms, each adding unique layers to linguistic complexity. These layers include structural variety, semantic richness, and cultural specificity. Structurally, idioms can range from simple phrases to complex sentences. Semantically, they carry meanings that cannot be deduced from the individual words alone. Culturally, idioms provide insights into the values, traditions, and humor of the people who use them.

To illustrate this diversity, consider the following types of idioms: irreversible binomials like 'bread and butter,' where the order of the words is fixed and conveys a specific meaning; reversible binomials such as 'on and off,' which can appear in either order without altering the meaning; and multino-

mials like 'Tom, Dick, and Harry,' which include a list of items or names. Additionally, there are phrasal verb idioms like 'give up,' where the meaning of the verb phrase is not easily inferred from its components. Moreover, cultural-pragmatic idioms like 'not my cup of tea' (meaning 'not really my thing') carry specific cultural connotations and are often context-dependent. For example, while 'not my cup of tea' is commonly used in British English, in Brazil, a similar sentiment is expressed with the idiom 'não é minha praia' (literally 'not my beach'). Recognizing and understanding these different types of idioms is crucial for grasping the full scope of idiomatic expressions in any language (Makkai, 1972).

Given this complexity, investigating idioms is particularly fascinating because they present significant challenges in second language acquisition. Idioms are often one of the last aspects of a language that are mastered, due to their figurative nature and cultural specificity (Irujo, 1986). Understanding idioms requires not only a good understanding of the language's vocabulary and grammar but also an appreciation of its cultural context and figurative nuances. This makes idioms a rich area of study for understanding the complexities of

language learning and cognitive processing in multilingual contexts.

Delving deeper into the intricacies of idiomatic expressions, it becomes apparent that their figurative meanings often rely on a property known as semantic decomposition (Nunberg, 1978). Semantic decomposition refers to the closeness of the connection between the literal meaning and the figurative meaning of an idiom. Specifically, it examines whether the constituent words of a semantically decomposable idiom contribute to the acquisition of its overall figurative meaning. In other words, when an idiom is semantically decomposable, its individual components can be linked to the idiomatic interpretation. This contrasts with nondecomposable idioms, where the overall meaning cannot be easily predicted from the literal meanings of its parts. For example, in the decomposable idiom 'spill the beans', the components can be connected to the figurative meaning of revealing a secret, whereas in the nondecomposable idiom 'kick the bucket', the literal components do not easily lead to the meaning of dying.

In adult discourse, idiomatic expressions are ubiquitous, seamlessly integrated into everyday speech to convey complex ideas and emotions efficiently. However, despite their prevalence in adult language, there remains a significant gap in understanding how children are exposed to and comprehend idioms. Investigating idiom comprehension in children not only sheds light on the developmental trajectory of language acquisition but also offers insights into the cognitive processes involved in understanding figurative language from an early age. Furthermore, studying idiom comprehension in children provides a unique opportunity to explore how language is acquired, processed, and integrated into broader linguistic and cognitive frameworks. By unraveling the mechanisms underlying children's understanding of idiomatic expressions, we gain a deeper understanding of the intricate interplay between language development, cognitive development, and cultural influences.

While research on language development has extensively explored children's acquisition of vocabulary, syntax, and pragmatic skills (Kim, 2017; Chomsky, 1969; Matthews, 2014), relatively little attention has been given to the specific process of idiom learning in childhood. Idioms pose a unique challenge to young language learners due to their

non-literal nature, requiring an understanding of context, metaphorical interpretation, and cultural knowledge beyond the literal meanings of individual words.

One of the few studies in this area is by Levorato & Cacciari (1992), which examined the role of context and familiarity in children's understanding of idioms. Their findings suggest that contextual cues and prior exposure to idiomatic language significantly influence children's comprehension of idioms, highlighting the importance of considering environmental factors in idiom acquisition. In addition, they found that children are able to understand the figurative meaning of idioms at about 9 years of age (Levorato & Cacciari, 1992).

Further research by Levorato et al. (2004) investigated the relationship between reading comprehension skills and idiom understanding in Italian children aged 7 to 10. Their study revealed that text comprehension skills strongly predicted idiom comprehension. Children with better reading comprehension could integrate figurative meanings with context more effectively. Over time, improvements in general comprehension skills — encompassing decoding ability, vocabulary knowledge, syntactic awareness, inferential skills, and contextual understanding — corresponded with enhanced idiom comprehension. This underscores the importance of inferential skills and contextual understanding in acquiring figurative language.

Expanding on the factors influencing idiom learning, research has also explored the decomposability of idioms and its impact on idiom comprehension. The study of Gibbs Jr. (1991) investigates the role of semantic decomposition — or semantic analyzability — in shaping children's understanding. By examining both decomposable (analyzable) and nondecomposable idioms, the research sheds light on the developmental trajectory of idiom comprehension. Younger children (6-7) tend to favor idioms with transparent meanings, while older children (8-11) demonstrate context-dependent understanding.

Furthermore, Caillies & Le Sourn-Bissaoui (2008) examined the link between idiom comprehension and theory of mind (ToM) in children aged 5 to 9. They discovered that higher ToM competences, particularly second-order ToM, correlated with better understanding of nondecomposable idioms. This suggests that children's ability to under-

stand others' mental states plays a crucial role in processing figurative language.

Building on the understanding of how cognitive abilities influence idiom comprehension, Sprenger et al. (2019) investigated the development of idiom knowledge across the lifespan, shedding light on age-related differences in idiom comprehension and production. Their study revealed a gradual improvement in idiom comprehension with age, suggesting that exposure and experience play crucial roles in the development of idiom proficiency.

Moreover, Ackerman (1982) explored how context and idiomatic form affect children's understanding of idioms. By presenting idioms within stories to children and adults, the study found that idiomatic interpretations were more frequent for standard idioms (those presented in their conventional form) than modified ones (those altered from their conventional form). Younger children (6-7 years old) tended to interpret idioms literally, while older children and adults showed improved idiomatic understanding, indicating a developmental progression in language processing.

The collective findings from these studies underscore the significant role that age and cognitive development play in idiom comprehension. For instance, Levorato et al. (2004) and Gibbs Jr. (1991) demonstrated that reading comprehension skills and the semantic decomposability of idioms significantly influence idiom comprehension. Furthermore, research by Caillies & Le Sourn-Bissaoui (2008) highlighted the importance of ToM in processing figurative language, particularly for nondecomposable idioms. Moreover, the studies of Ackerman (1982) and Sprenger et al. (2019) illustrate the developmental progression in idiom comprehension, emphasizing the influence of age and experience on language processing abilities.

Despite these advances in understanding the developmental aspects of idiom comprehension, there remains a need to delve deeper into the specific idioms children are exposed to in their natural linguistic environments and how they interpret these expressions. Identifying the types and frequency of idioms encountered by children can provide a clearer picture of their idiom acquisition process, thereby enriching our understanding of how idiomatic knowledge is built from a young age.

Arbeek (2024) conducted a corpus study on idioms in Dutch children's books, producing a com-

prehensive list of every idiom found in the corpus along with their respective frequencies. This research was undertaken to address the gap in understanding how often and in what contexts children encounter idiomatic expressions in their reading materials. By identifying the frequency and variety of idioms in child-directed texts, Arbeek aimed to provide a foundational understanding of the linguistic environment children are exposed to. However, a crucial question remains: do children actually know and understand these idioms?

To address this question, the present study aims to fill the gap by investigating whether children aged 7 to 11 can demonstrate familiarity with idiomatic expressions identified in the corpus study of Arbeek (2024). By conducting a multiple-choice questionnaire on Dutch children, this research will explore the extent to which children recognize these idioms. This is important for enhancing our theoretical understanding of language acquisition and cognitive development. Ultimately, this study seeks to contribute to a more nuanced understanding of idiom learning in children, bridging the gap between exposure and comprehension in the context of their linguistic environment.

This study will specifically examine the influence of exposure and age on idiom familiarity in children. The primary research questions guiding this study are: (1) To what extent does exposure to idioms in children's literature influence idiom familiarity among Dutch children? and (2) How does age affect idiom familiarity in children aged 7 to 11? The hypothesis is that (1) idioms with higher frequency in children's books will be rated familiar more often by the children. Additionally, it is hypothesized that (2) older children will demonstrate greater familiarity with idiomatic expressions than younger children, reflecting developmental progress in language comprehension. Through this investigation, the present study aims to provide valuable insights into the stages of idiom acquisition and the factors that influence children's ability to grasp figurative language.

## 2 Methods

This study investigates the influence of exposure and age on idiom familiarity in Dutch children aged 7 to 11. The experiment involved administering a multiple-choice questionnaire to children in different school classes to assess their recognition and familiarity with idiomatic expressions identified in the corpus study of Arbeek (2024). The methodology encompasses participant recruitment, experimental procedure, and the materials and design used to ensure accurate and meaningful data collection.

### 2.1 Participants

Participants were recruited by contacting Dutch primary schools through email. A total of 44 children aged 7-11 participated in this study, with a mean age of 8.8 (SD = 0.95). Three school classes contributed to the experiments. The first experiment was in year 6 of the school *Widar Vrijeschool* in the region of Groningen, in the North of the Netherlands. From this class, 20 participants made the questionnaire. This class had a mean age of 9.5 (SD = 0.59). The second experiment was in year 4 of the same school (*Widar Vrijeschool*), where 17 participants made the questionnaire. This class had a mean age of 7.8 (SD = 0.38). The third experiment was in year 6 of the school *Hart van Slangenbeek* in the region of Twente, in the East of the Netherlands. From this class, 7 participants made the questionnaire. This class had a mean age of 9.4 (SD = 0.50).

Participants did not receive compensation for their participation. Moreover, the research was conducted as part of a normal school day, integrating smoothly into the students' regular classroom activities.

### 2.2 Procedure

The participants were informed about the experiment a few weeks in advance. They were told that the experiment involved filling out a short questionnaire and assured that it would not affect their school performances. Prior to conducting the study, ethical approval was obtained from the Research Ethics Committee (CETO) at the University of Groningen. The committee reviewed the research

plan, provided feedback for adjustments to the informed consent form, and subsequently granted approval. The parents of the participants were required to sign informed consent forms for their child(ren)'s participation. For two of the classes, the consent forms were not returned by the parents. However, the teacher from *Widar Vrijeschool*, who managed the consent forms and informed the participants, declared that all parents had consented and signed a statement to this effect.

The experiments took place in the classrooms of the participants, during school time, to ensure a familiar and comfortable environment for the children, thereby reducing potential anxiety and enhancing their ability to focus on the task. Before handing out the questionnaires, participants were provided with some information on the purpose of the study by the researcher, emphasizing that it is not about performance, grading, or right or wrong answers. Additionally, instructions for filling in the questionnaire were given by the researcher, which were also provided at the top of the questionnaire. This included guidance on the nature of the questions, the importance of honest responses, and encouragement to take their time and ask for help if needed. For instance, participants were encouraged to ask for clarification if they did not understand a question.

Before the general questions started, participants were asked to fill in their age. No other personally identifiable data was collected during this study. The questionnaires were completed individually but simultaneously, in silence, to ensure independent and unbiased responses while maintaining a controlled environment. There was no time limit for the questionnaire, to allow participants to think carefully about their answers and to minimize stress or pressure. Furthermore, teachers were given the flexibility to integrate the experiment into their lessons in their own way, which helped ensure the experiment was smoothly incorporated into the existing classroom schedule. Additionally, teachers were responsible for determining what the children should do once they had completed the questionnaire, ensuring a seamless transition back to regular classroom activities and maintaining classroom order. For example, two teachers allowed students to read quietly, while the other teacher permitted them to engage in independent or group projects.

## 2.3 Materials and Design

Figure 2.1 shows the idioms used in the questionnaires. The list of idioms consists of 26 Dutch idioms selected from the list composed by Arbeek (2024), who conducted a corpus study on idioms in Dutch children’s books to identify the frequency and variety of idioms encountered by children. The selection process for this study was carried out in several steps to ensure a representative and manageable set of idioms.

First, the complete list was divided into five groups based on frequency distribution: the first group ranges from a frequency of 251 to 79 (only 5 idioms fall within this range), the second group from 32 to 23 (there are no idioms with a frequency between 79 and 32), the third from 18 to 12, the fourth from 11 to 7, and the fifth from 6 to 3. This division aimed to cover a broad spectrum of idioms from those encountered rarely to those seen frequently, providing a comprehensive analysis of idiom familiarity across different exposure levels.

From each of these frequency-based groups, five idioms were randomly selected, resulting in a preliminary list of 25 idioms. This means that all five idioms from the first group were selected. To achieve an even number of idioms, one additional idiom was randomly selected from the entire list (*‘Aan de slag gaan’*, from the second group), resulting in a total of 26 idioms. Figure 2.1 shows the final set of idioms, ordered on frequency.

This final set was then divided into two lists of 13 idioms each, with the idioms alternated between the lists to maintain balance in frequency distributions. This division was designed to prevent the questionnaire from becoming too lengthy and to make it more manageable for the participants, particularly because they are children. Research has shown that shorter tasks are more effective for maintaining the attention and engagement of young children during cognitive assessments (Mari, 2024).

Lastly, the order of the lists were shuffled to end up with 4 versions of the questionnaire; A1, A2, B1 and B2. The letter denotes the idiom list used in the questionnaire, whereas the number indicates the order in which the questions appear. This shuffling was implemented to control for potential order effects, which can influence participants’ responses based on the sequence of the questions presented (Shaughnessy et al., 2006). Each question-

Frequencies of Idioms (Arbeek, 2024)



Figure 2.1: The idioms used in this study, with their frequencies as found by Arbeek (2024). The numbers behind the idioms indicate their positions in the complete list of Arbeek (2024).

naire consisted of 13 idioms, and every participant completed only one version of the questionnaire. However, the version is not indicated on the questionnaires to prevent any potential bias that might arise if participants are aware of the specific version they are completing. Ensuring participants are unaware of the questionnaire version helps maintain the integrity of their responses and reduces the likelihood of any preconceived notions influencing their answers (Krosnick & Berent, 1993).

The questionnaire consisted of multiple-choice questions designed to assess the participants’ familiarity with various idioms. Each question presented one idiom, written out in full, followed by a prompt asking the participant to indicate how often they have encountered the idiom. The question number, denoted as ‘Vraag n’, where ‘n’ represents the specific question number, preceded each idiom.

The questions are structured as follows:

Vraag  $n$  (*Question n*): 'Idiom'

Hoe vaak heb jij deze uitdrukking gezien of gehoord?  
(*How often have you seen or heard this idiom?*)

The participants get three answer options; 'Nooit' (*Never*), 'Soms' (*Sometimes*), 'Vaak' (*Often*). The participants are instructed to pick one of the options. This design was chosen based on previous research methodologies, such as those used by Levorato & Cacciari (1992), which employed a multiple-choice questionnaire with three answer options to assess idiom comprehension. Additionally, Levorato et al. (2004) utilized a scaling approach to categorize children's idiom comprehension into Poor, Medium, and Good levels. These precedents support the effectiveness of multiple-choice formats in evaluating figurative language understanding in children, ensuring clarity and ease of response for the participants. The questionnaires were printed and made on paper to ensure familiarity, accessibility, and ease of interaction with the materials. A full questionnaire can be found in Appendix A.

In summary, this study employed a robust methodology to investigate the influence of exposure and age on idiom familiarity among Dutch children aged 7 to 11. Through careful participant recruitment, ethical considerations, and a structured experimental procedure integrated into the school day, the study aimed to provide comprehensive insights into children's familiarity with Dutch idiomatic expressions. The questionnaire design, based on previous research methodologies (Levorato & Cacciari, 1992; Levorato et al., 2004), ensured clarity and accessibility for young participants. By utilizing a controlled environment and balanced questionnaire versions, the study aimed to mitigate potential biases and order effects. The expected outcomes include insights into how exposure and age impact idiom familiarity among children, contributing to a deeper understanding of language acquisition and cognitive development in educational settings.

## 3 Results

As mentioned in the previous section, a total of 44 children aged 7-11 participated in the experiment. However, the 7 year olds (3 participants) and 11 year olds (2 participants) were excluded from data analysis, since they did not cover all versions of the questionnaire. Including these participants would give misleading results due to the different idioms used in the A and B versions. There were 2 participants with reading difficulties. The questionnaire was read out loud to these participants. These participants were not excluded from data analysis, because the teacher was convinced this would not have influenced the given answers. Ultimately, 39 participants were included in the analysis, and the mean age in the analyzed data is 8.925. The decision to exclude certain age groups was made post hoc and was necessary to maintain the integrity of the results. Otherwise, all participants would have been included.

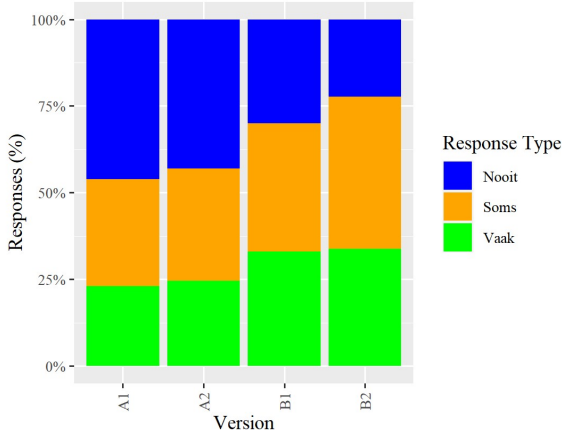
### 3.1 General Analysis

This section presents an overview of the response distributions and general familiarity with idioms among the participants. Figure 3.1 shows the response distribution for each version of the questionnaire. Versions A1 and A2 show similar responses, with response type 'vaak' given just under 25% of the time. 'soms' is responded 31% in version A1 and 32% in version A2. Version B1 has more 'vaak' responses (33%) and more 'soms' responses (37%) than A1 and A2. Version B2 has a similar amount of 'vaak' responses (34%), but more 'soms' responses (44%) than version B1.

In short, the idioms from the B-versions generated more familiarity responses ('vaak' and 'soms') than the idioms from the A-versions. To statistically validate this observation, a two-proportion z-test was conducted comparing the pooled responses of versions A (A1 and A2 combined) and versions B (B1 and B2 combined). Results show that versions B elicited a significantly higher proportion of familiarity responses compared to versions A ( $X^2 = 18.58$ ,  $p < 0.05$ ).

The data also shows that, on average, each student knew approximately 8.2 idioms (in one questionnaire, so out of 13 idioms), with a standard deviation of 2.6 idioms. The overall familiarity re-

**Response Distribution by Version**



**Figure 3.1: Response distribution for each version of the questionnaire.**

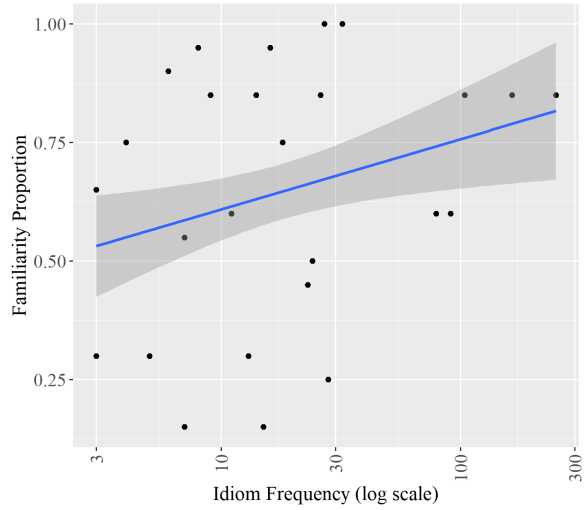
sponses across all idioms were 64.6%, indicating the level of idiom familiarity among the participants.

### 3.2 Frequency

The first question we want to answer is how the frequency of idioms relates to the proportion of familiarity among the participants. Figure 3.2 shows the relationship between the idiom frequency, as found in the study of Arbeek (2024), and the familiarity proportion. In this plot, idiom frequency on the x-axis is presented in a log scale to enhance data interpretability and to better visualize the spread of data across different frequencies.

The familiarity proportion is calculated by dividing the sum of 'soms' and 'vaak' responses by the total number of responses. To illustrate, a low frequency idiom 'Lachen als een boer met kiespijn' (*To laugh like a farmer with a toothache*, meaning to laugh insincerely) has a frequency of 5 and a familiarity proportion of 0.30, and a high frequency idiom like 'In de gaten hebben' (*To have in the holes*, meaning to keep an eye on or to be aware of something) has a frequency of 164 and a familiarity proportion of 0.85. The plot depicts a large variance for data points with a frequency of 30 and below, which is where the majority of data points are located. For example, the idiom 'Een brok in zijn keel voelen' (*To feel a chunk in his throat*, meaning to be moved, emotionally) with a frequency of 14, has a

**Relationship Frequency and Familiarity Proportion**



**Figure 3.2: The relationship between idiom frequencies as found by Arbeek (2024) and familiarity proportion. The x-axis is converted to a log scale to enhance interpretability.**

familiarity proportion of 0.85, while 'Zijn ogen ten hemel slaan' (*To turn his eyes to heaven*, meaning to look up to the sky) with a frequency of 15, has a proportion of 0.15.

The plot can be conceptually divided into four quadrants: the top-left quadrant with high familiarity proportion and low idiom frequency; the top-right quadrant with high familiarity proportion and high idiom frequency; the bottom-left quadrant with low familiarity proportion and low idiom frequency; and the bottom-right quadrant with low familiarity proportion and high idiom frequency. Notably, there is an empty region in the bottom-right quadrant, indicating no data points are located in the higher range of idiom frequency and the lower range of familiarity proportion.

A simple linear regression analysis was conducted to explore the relationship between the logarithm of frequency and familiarity proportion. The regression model was statistically significant ( $F(1, 76) = 6.46, p < 0.05$ ), with an R-squared value of 0.07834. This indicates that 7.8% of the variance in the percentage known is explained by the model, which is relatively small. Generally, an R-squared value above 0.5 is considered moderate, while val-

ues above 0.7 are seen as strong indicators of a model’s explanatory power (Frost, 2023; Muralidhar, 2023). The low R-squared value in this case suggests that other factors beyond idiom frequency likely play a role in determining familiarity proportions among the participants.

Both the intercept ( $\beta = 0.45$ ,  $p < 0.05$ ) and the slope ( $\beta = 0.06$ ,  $p < 0.05$ ) were statistically significant. These results suggest a positive relationship between the logarithm of frequency and the percentage of people who know the respective idiom, with higher frequencies corresponding to higher familiarity percentages.

Even though converting the x-axis to a log scale — to enhance data interpretability — influences the linear regression analysis, the same analysis on the real scale provides significant results as well ( $F(1, 76) = 6.565$ ,  $p < 0.05$ ). In short, there seems to be a significant positive relationship between idiom frequency (as found by Arbeek (2024)) and familiarity proportion.

However, the assumption of homoscedasticity appears questionable given the large variance in data points with frequencies below 30, and the little variance in data points with higher frequencies. Homoscedasticity (a constant variance of the residuals) is important because it ensures that the regression model provides unbiased and efficient estimates of the coefficients, and that the standard errors of these estimates are reliable. When this assumption is violated, it can lead to inefficient estimates and affect the validity of hypothesis tests (Williams, 2015). To formally test for heteroscedasticity in the present study, a Breusch-Pagan test was conducted. The test revealed significant heteroscedasticity (BP = 5.29,  $p < 0.05$ ), confirming the initial suspicion.

To account for this heteroscedasticity, robust standard errors were calculated using a Heteroscedasticity-Consistent Covariance Matrix Estimator (HC). Specifically, the HC3 adjustment was used, since this adjustment is particularly useful for small samples. This approach ensures more reliable estimates of the standard errors, p-values, and confidence intervals (Long & Ervin, 2000). The results of the coefficient test with robust standard errors confirmed the statistical significance of both the intercept ( $\beta = 0.46$ ,  $p < 0.05$ ) and the slope ( $\beta = 0.06$ ,  $p < 0.05$ ). These findings indicate that the positive relationship between the logarithm of fre-

quency and familiarity proportion holds even after adjusting for heteroscedasticity.

In short, while the presence of heteroscedasticity necessitated adjustments to the statistical tests, the overall findings remain robust. However, it should be kept in mind that other factors beyond idiom frequency likely influence idiom familiarity.

### 3.3 Age

The second question we want to answer is how idiom familiarity develops through different ages. An overview of the results per age can be found in Appendix B. It shows the percentages of given responses per idiom, for each age group. The idioms are ordered on the percentage of familiarity responses (‘vaak’ and ‘soms’). In every age group there are idioms with low familiarity and idioms with high familiarity. However, the order of the idioms differs between age groups.

To measure familiarity, the previously used calculation in section 3.2 was used to get a familiarity proportion (F):

$$F = \frac{\text{'soms' responses} + \text{'vaak' responses}}{\text{total responses}}$$

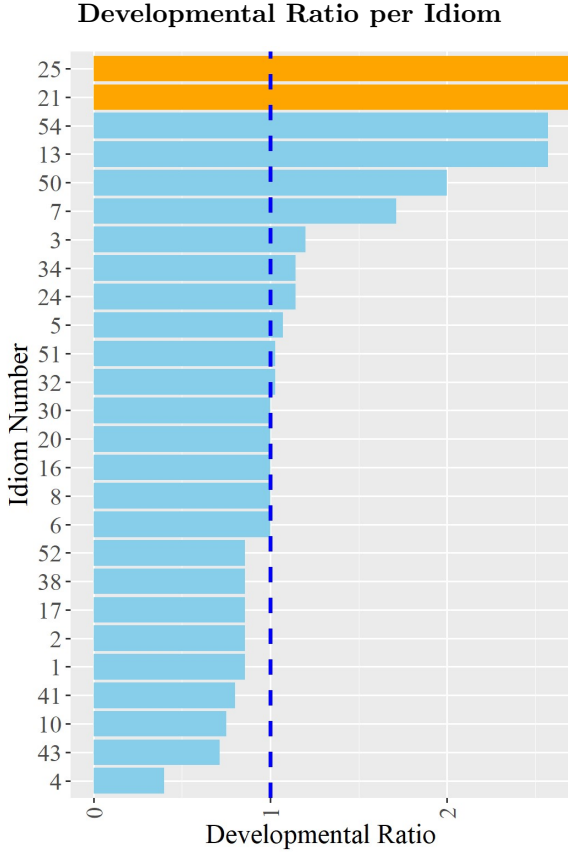
This proportion provides a measure of how frequently participants are familiar with the idioms in question. Table 3.1 summarizes the familiarity proportions for different age groups, demonstrating how familiarity varies across these groups. Familiarity proportions of age 8 and 10 are the same ( $F_8 = 0.66$ ,  $F_{10} = 0.66$ ), whereas the familiarity proportion of age 9 is slightly lower ( $F_9 = 0.62$ ).

<b>Familiarity<sub>Age</sub></b>	<b>Value</b>
$F_8$	0.66
$F_9$	0.62
$F_{10}$	0.66

**Table 3.1: Familiarity proportions for each age group.**

A Kruskal-Wallis test was conducted to examine the differences in familiarity proportions among the age groups. The test revealed no significant difference in familiarity proportions across age groups





**Figure 3.3:** Developmental ratios for each idiom, with idioms 25 and 21 highlighted in orange due to their infinite ratios. The dashed blue line at a ratio of 1 indicates the threshold where ratios shift from decrease to increase.

( $H = 3.01, p > 0.05$ ). Given the non-significant result of the Kruskal-Wallis test, no post hoc test was deemed necessary.

Familiarity trends with individual idioms across different ages were analyzed using a metric referred to as the ‘developmental ratio’ ( $D_{idiom}$ ). This ratio shows the trend in which familiarity has increased or decreased with older age. These ratios were calculated by dividing the familiarity proportion of each idiom for age 10 by the familiarity proportion of the same idiom for age 8:

$$D_{idiom} = \frac{F_{10,idiom}}{F_{8,idiom}}$$

This method provides insight into how much the

familiarity of each idiom increased or decreased relative to age 8. This suggests an idiom with a developmental ratio of 1 is as familiar to participants aged 8 as to participants aged 10. Figure 3.3 shows the developmental ratio for all individual idioms. The ratios range between 0.4 (‘In de rede vallen’ (*To fall in the speech*, meaning to interrupt)) and 2.57 (‘Voor de hand liggen’ (*To lay in front of the hand*, meaning to be obvious) and ‘Lachen als een boer met kiespijn’ (*To laugh like a farmer with a toothache*, meaning to laugh insincerely)). The two ratios of ‘Zijn ogen ten hemel slaan’ (*To turn his eyes to heaven*, meaning to look up to the sky) and ‘Het niet over zijn hart verkrijgen’ (*To not get it over his heart*, meaning to be unable to do something, with emotional reasons) were infinite because the familiarity at age 8 was zero, so any increase in familiarity at age 10 results in an infinite developmental ratio. These idioms are marked orange in figure 3.3. Overall, 12 idioms showed an increase in familiarity ( $D_{idiom} > 1$ ) and 9 idioms showed a decrease in familiarity ( $D_{idiom} < 1$ ). A complete list of the developmental ratios is presented in Appendix C.

To look for possible differences with developmental ratios in other age ranges, the ratios were also calculated for the original data (i.e. development in familiarity between participants aged 7 and 11). However, due to the small number of participants, the ratio values of the idioms were either infinite, 2, 1 or non-existent. A comparison would therefore be inconclusive and not very meaningful.

There are idioms that stand out in deeper data analysis. For instance, idiom number 6 — ‘Van top tot teen’ (*From top to toe*, meaning from head to toe) — and idiom number 8 — ‘Aan de slag gaan’ (*To go on the move*, meaning to get started) — have a developmental ratio of 1. At first sight, this does not seem very interesting. However, when interpreting the figure in Appendix B, it appears that these two idioms are the only idioms known by every participant in every age group.

## 4 Discussion

This study aimed to investigate whether children aged 7 to 11 demonstrate familiarity with idiomatic expressions found in Dutch children’s books, as identified in the corpus study of Arbeek (2024). A

multiple-choice questionnaire was administered to a sample of 44 children. Data analysis was limited to participants aged 8 to 10 due to incomplete coverage of the questionnaire versions by the youngest and oldest participants. This section will interpret the results detailed earlier, highlighting their significance in relation to the research objectives and relevant literature. Additionally, it will critically evaluate the study's limitations and offer recommendations for future research directions in this field.

## 4.1 Frequency

Investigating the relationship between idiom frequency and familiarity is crucial for understanding how children acquire figurative language. Previous studies, such as those by Levorato & Cacciari (1992) and Levorato et al. (2004), have highlighted the importance of contextual cues and prior exposure in children's idiom comprehension. These studies suggest that frequent encounters with idiomatic expressions in reading materials significantly influence children's understanding and familiarity with idioms. In light of these findings, our study aimed to explore whether the frequency of idioms in children's literature correlates with familiarity among young readers.

The linear regression analysis revealed a statistically significant positive relationship between idiom frequency (as found by Arbeek (2024)) and the proportion of participants familiar with the idioms. Specifically, higher frequencies of idioms in children's books corresponded to higher knowledge proportions among the participants. This finding aligns with the conclusions of Levorato et al. (2004), who emphasized the importance of exposure and reading comprehension skills in idiom understanding. The present study extends these insights by demonstrating that frequent exposure to idioms in reading materials is associated with greater familiarity among children.

However, the scatter plot analysis revealed some peculiarities in the data. For instance, certain idioms with moderate to low frequency in the corpus were unexpectedly well-known among the children. An example of such an idiom is 'Een oogje in het zeil houden' ('*To hold an eye in the sail*', meaning to keep an eye on things). This idiom has a frequency of 8, but a familiarity proportion of 0.95.

This anomaly might be explained by the influ-

ence of other sources, such as teachers using these idioms frequently in educational settings. For example, the idiom 'Aan de slag gaan' ('*To go on the move*', meaning to get started) is commonly used by teachers to tell children to begin their tasks. Additionally, children might encounter these idioms in other media outside of books, such as television or the internet. Moreover, (grand)parents or guardians might use specific idioms in their conversations at home, contributing to the children's familiarity with them. These additional sources of exposure can create noise in the data because the actual frequency of exposure is higher than what is recorded in the corpus. If more accurate information on all sources of idiom exposure were available, it could reduce this noise by providing a clearer picture of the true frequency with which children encounter each idiom. Furthermore, the decomposability of idioms could also play a role; idioms that are more semantically transparent might be easier for children to understand and remember, regardless of their frequency in the corpus. Decomposability will be further discussed in section 4.3.

## 4.2 Age

Investigating age-related differences in idiom familiarity is important because it can provide insights into the cognitive and linguistic development. Previous research by Ackerman (1982) and Sprenger et al. (2019) suggested that idiom comprehension improves with age, showing a developmental trajectory where older children demonstrate better understanding of idioms than younger ones. Contrary to these findings, our analysis of age-related differences indicated no significant variability in idiom familiarity across different age groups. While certain idioms, such as 'Van top tot teen' ('*From top to toe*', meaning from head to toe) and 'Aan de slag gaan' ('*To go on the move*', meaning to get started) were universally recognized, the familiarity with other idioms varied among different ages. The familiarity ratios showed no significant increase in familiarity with increased age. This finding challenges the notion of a clear developmental progression in idiom comprehension.

Another interesting observation was the familiarity ratio for specific idioms. For example, idioms like 'Voor de hand liggen' ('*To lay in front of the hand*', meaning to be obvious) and 'Lachen als een

boer met kiespijn' (*To laugh like a farmer with a toothache*, meaning to laugh insincerely) showed substantial increases in familiarity ratios between ages 8 and 10. On the other hand, idioms such as 'Zijn hart klopt in zijn keel' (*His heart beats in his throat*, meaning to have fear) and 'Zijn woorden kracht bij zettten' (*To add strength to his words*, meaning to emphasize what he is saying) showed a decrease in familiarity from age 8 to age 10. Even the idiom with the highest frequency from the corpus — 'Iets aan de hand zijn' (*To be something on the hand*, meaning that something is the matter) — showed a decrease in developmental familiarity.

Possible reasons for these variations could include differential exposure to certain idioms in educational or home environments, or specific teaching practices that emphasize certain idiomatic expressions over others. Specific age-related differences may be influenced by the varying cognitive and linguistic development stages of children. Younger children might have more frequent and diverse exposure to certain idioms through children's literature or educational content designed for their age group, while older children might encounter a different set of idiomatic expressions that are more context-specific or abstract. This differential exposure can result in variations in familiarity ratios. Moreover, as children grow older, their interests and the media they consume might change, potentially leading to less frequent encounters with idioms that were more common in their earlier years. Social and cultural factors, such as peer influence and the language used in different social settings, can also play a role in how idiom familiarity develops with age.

### 4.3 Semantic Decomposition

The concept of semantic decomposition, as discussed by Gibbs Jr. (1991), also plays a crucial role in idiom comprehension. They suggest that children are more likely to understand idioms that are semantically decomposable, where the figurative meaning can be inferred from the literal meanings of the components. This is particularly relevant for younger children, who tend to favor idioms with transparent meanings. This is consistent with the work of Caillies & Le Sourn-Bissaoui (2008),

who found that higher theory of mind (ToM) competences correlated with better understanding of nondecomposable idioms. Thus, children's cognitive abilities, including ToM, are integral to processing more complex, nondecomposable idioms.

There is no comprehensive official measurement for the decomposability of all idioms included in this study. Despite this, it is likely that some idioms in the list exhibit higher levels of semantic decomposability than others. For instance, idioms like 'Van top tot teen' (from head to toe) — which was rated familiar by every participant — might be more semantically transparent compared to others, allowing children to infer their figurative meanings more easily. However, it should be noted that even some supposedly nondecomposable idioms were well understood, such as 'In de gaten hebben' (*To have in the holes*, meaning to keep an eye on or to be aware of something), which has a familiarity proportion of 0.85. This is likely due to their high frequency and contextual exposure in reading materials and educational settings.

The variation in decomposability among these idioms suggests that semantic decomposition can play a significant role in idiom comprehension, as opposed to relying solely on frequency of exposure. This highlights the need for further research to systematically measure the decomposability of idioms and examine how this factor, compared to others like frequency of exposure, influences children's familiarity and comprehension.

### 4.4 Limitations

While this study provides valuable insights into children's idiom familiarity, several limitations are at hand. First, expanding the sample size and including a more balanced representation of all age groups would allow for a more comprehensive analysis of age-related differences.

Second, the use of four different versions might give a fragmented look of the results. The youngest and oldest participants were excluded from the analysis due to incomplete coverage of the idioms across the different questionnaire versions. Moreover, the performance on versions A compared to versions B is skewed, with versions B significantly showing more familiarity responses. This disparity suggests potential inconsistencies in the difficulty or decomposability of the idioms presented in differ-

ent versions. Instead of alternating idioms for different versions, more careful consideration should be given to the selection of idioms to ensure balanced difficulty and familiarity across all versions. In addition, future studies should ensure that all age groups are adequately represented across all versions of the questionnaire to allow for more robust comparisons.

Third, the idiom list used in this study, while comprehensive, may not fully capture the range of idiomatic expressions children are exposed to in their daily lives. Including idioms from a wider array of sources, such as spoken language and other media, could provide a more holistic view of children's idiom familiarity.

Lastly, the geographic location of the participants and their schools could have influenced the results. Cultural and regional variations in language use may affect the familiarity and comprehension of certain idioms. Future research should consider including participants from diverse geographic backgrounds to examine these potential differences.

## 4.5 Future Research

As discussed in the previous sections, future research should address several key areas to build on the findings of this study. Increasing the sample size and ensuring a balanced representation across all age groups will provide a more reliable analysis of age-related differences in idiom familiarity. Also, harmonizing the questionnaire versions by carefully selecting idioms of consistent difficulty and familiarity can help avoid disparities and ensure more reliable results. Additionally, expanding the idiom list to include expressions from a broader range of sources, such as spoken language and other media, will offer a more complete understanding of children's idiom familiarity. Moreover, incorporating participants from diverse geographic and cultural backgrounds will help examine regional variations in language use and their impact on idiom comprehension. Finally, systematically measuring the decomposability of idioms will elucidate the role of semantic transparency in children's idiom understanding, providing deeper insights into cognitive and developmental factors influencing figurative language acquisition.

## 4.6 Conclusion

This study aimed to investigate the influence of exposure and age on idiom familiarity among children aged 7 to 11. The research questions were: (1) To what extent does exposure to idioms in children's literature influence idiom familiarity among Dutch children? and (2) How does age affect idiom familiarity in children aged 7 to 11? The hypotheses were that (1) idioms with higher frequency in children's books would be rated as familiar more often by the children and that (2) older children would demonstrate greater familiarity with idiomatic expressions than younger children.

Our findings confirmed the first hypothesis: there is a significant positive relationship between the frequency of idioms in children's literature and the proportion of children familiar with those idioms. This suggests that frequent exposure to idiomatic expressions in reading materials substantially enhances children's familiarity with these expressions, supporting the importance of incorporating idioms into educational content to aid language development.

As for the second hypothesis, the study found no significant age-related differences in idiom familiarity across the ages of 8 to 10. This result challenges the assumption of a straightforward developmental progression in idiom comprehension, indicating that factors other than age, such as the context in which idioms are encountered or their semantic decomposability, may play a more critical role in how children understand and retain idiomatic expressions.

In conclusion, this study highlights the impact of idiom frequency on children's familiarity with idiomatic expressions, emphasizing the value of frequent exposure through reading materials. Age did not emerge as a significant factor for idiom familiarity in this study. Future research should expand sample sizes, harmonize idiom selection across questionnaire versions, measure semantic decomposition and include a more diverse participant pool to further elucidate the factors influencing idiom familiarity and comprehension among children.

## References

- Ackerman, B. P. (1982). On comprehending idioms: Do children get the picture? *Journal of Experimental Child Psychology*, 33(3), 439-454.
- Arbeek, J. (2024). *Investigating dutch children's idiom exposure*. (Unpublished bachelor's thesis)
- Caillies, S., & Le Sourn-Bissaoui, S. (2008). Children's understanding of idioms and theory of mind development. *Developmental Science*, 11(5), 703-711.
- Chomsky, C. (1969). The acquisition of syntax in children from 5 to 10. *M.I.T. Press*.
- Frost, J. (2023). *Interpreting r-squared (coefficient of determination) for linear regression*. Retrieved from <https://statisticsbyjim.com/regression/interpret-r-squared-regression/>
- Gibbs Jr., R. W. (1991). Semantic analyzability in children's understanding of idioms. *Journal of Speech, Language, and Hearing Research*, 34(3), 613-620.
- Irujo, S. (1986). Don't put your leg in your mouth: Transfer in the acquisition of idioms in a second language. *tesol Quarterly*, 20(2), 287-304.
- Kim, Y.-S. G. (2017). Multicomponent view of vocabulary acquisition: An investigation with primary grade children. *Journal of Experimental Child Psychology*, 162, 120-133.
- Krosnick, J. A., & Berent, M. K. (1993). Comparisons of party identification and policy preferences: The impact of survey question format. *American Journal of Political Science*, 941-964.
- Levorato, M. C., & Cacciari, C. (1992). Children's comprehension and production of idioms: the role of context and familiarity. *Journal of Child Language*, 19(2), 415-433.
- Levorato, M. C., Nesi, B., & Cacciari, C. (2004). Reading comprehension and understanding idiomatic expressions: A developmental study. *Brain and Language*, 91(3), 303-314.
- Long, J. S., & Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician*, 54(3), 217-224.
- Makkai, A. (1972). *Idiom structure in english*. Berlin, Boston: De Gruyter Mouton.
- Mari, A. (2024). Making conventional data collection more child-friendly: Questionnaires with young students. *African Evaluation Journal*, 12(1), 690.
- Matthews, D. (2014). *Pragmatic development in first language acquisition*. John Benjamins Publishing Company.
- Muralidhar, K. (2023, February 8). *Demystifying r-squared and adjusted r-squared*. Retrieved from <https://builtin.com/data-science>
- Nunberg, G. D. (1978). *The pragmatics of reference*. City University of New York.
- Oxford English Dictionary. (2023). *Idiom, n*. Oxford University Press. Retrieved from <https://doi.org/10.1093/OED/5576832890>
- Shaughnessy, J., Zechmeister, E., & Zechmeister, J. (2006). *Research methods in psychology*. McGraw-Hill Higher Education.
- Sprenger, S. A., la Roi, A., & Van Rij, J. (2019). The development of idiom knowledge across the lifespan. *Frontiers in Communication*, 4, 29.
- Williams, R. (2015). Heteroscedasticity. *Journal of Research Methods and Statistics*, 7(2), 345-362.

## A Questionnaire Version A1

### Vragenlijst Nederlandse Uitdrukkingen

In deze vragenlijst zul je een aantal uitdrukkingen lezen. Het is aan jou om aan te geven hoe vaak je deze uitdrukking hebt gehoord of gezien (bijvoorbeeld omdat je vader/moeder/leraar het heeft gezegd of als je het in een boek hebt gelezen). Er zijn geen goede of foute antwoorden, dus kies gewoon wat jij denkt dat het beste is!

Het is belangrijk dat je eerlijk antwoord geeft op elke vraag. Er zijn geen trucjes en we willen graag weten wat jij echt denkt. Neem je tijd om elke vraag te lezen en kies het antwoord dat het beste bij jou past.

Als je een vraag niet begrijpt, vraag dan gerust om hulp.

Bedankt voor het meedoen!

---

Hoe oud ben je? \_\_\_\_\_

Vraag 1: "**er nog een schepje bovenop doen**"

Hoe vaak heb jij deze uitdrukking gezien of gehoord?

- Nooit
- Soms
- Vaak

Vraag 2: "**het niet over zijn hart verkrijgen**"

Hoe vaak heb jij deze uitdrukking gezien of gehoord?

- Nooit
- Soms
- Vaak

Vraag 3: "**hoe de vork in de steel zit**"

Hoe vaak heb jij deze uitdrukking gezien of gehoord?

- Nooit
- Soms
- Vaak

Vraag 4: "**in de steek laten**"

Hoe vaak heb jij deze uitdrukking gezien of gehoord?

- Nooit
- Soms
- Vaak

Vraag 5: "**voor de hand liggen**"

Hoe vaak heb jij deze uitdrukking gezien of gehoord?

- Nooit
- Soms
- Vaak

Vraag 6: "**de stuipen op het lijf jagen**"

Hoe vaak heb jij deze uitdrukking gezien of gehoord?

- Nooit
- Soms
- Vaak

Vraag 7: "**te mooi om waar te zijn**"

Hoe vaak heb jij deze uitdrukking gezien of gehoord?

- Nooit
- Soms
- Vaak

Vraag 8: "**zijn ogen ten hemel slaan**"

Hoe vaak heb jij deze uitdrukking gezien of gehoord?

- Nooit
- Soms
- Vaak

Vraag 9: "**per slot van rekening**"

Hoe vaak heb jij deze uitdrukking gezien of gehoord?

- Nooit
- Soms
- Vaak

Vraag 10: "**op het punt staan om**"

Hoe vaak heb jij deze uitdrukking gezien of gehoord?

- Nooit
- Soms
- Vaak



Vraag 11: "**het einde van de wereld**"

Hoe vaak heb jij deze uitdrukking gezien of gehoord?

- Nooit
- Soms
- Vaak

Vraag 12: "**iets aan de hand zijn**"

Hoe vaak heb jij deze uitdrukking gezien of gehoord?

- Nooit
- Soms
- Vaak

Vraag 13: "**zijn woorden kracht bij zetten**"

Hoe vaak heb jij deze uitdrukking gezien of gehoord?

- Nooit
- Soms
- Vaak

---

Dit is het einde van de vragenlijst. Bedankt voor het invullen!

## B Responses per Age

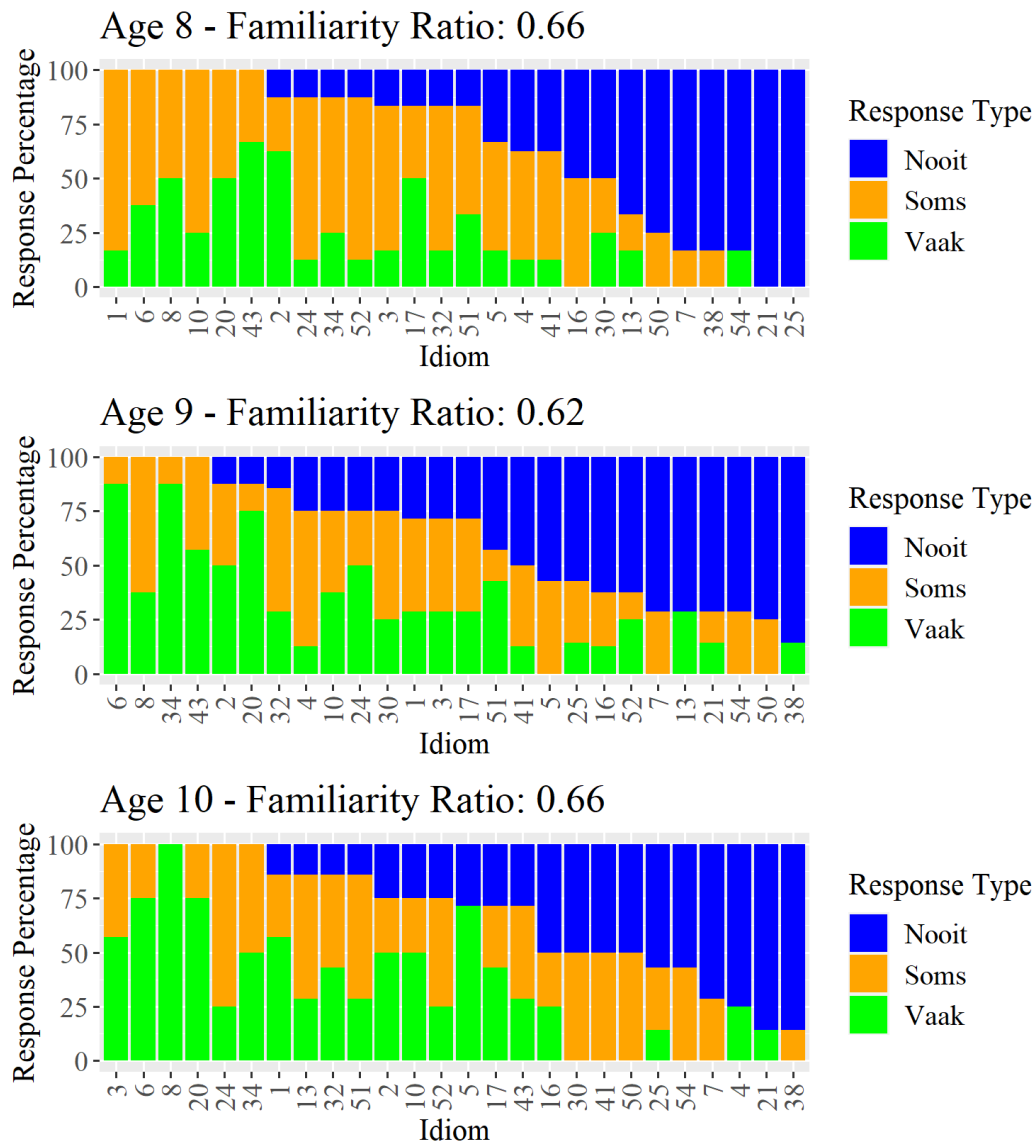


Figure B.1: Response percentages for each age group and idiom. Each subplot is ordered on familiarity proportion, with the overall familiarity proportion shown in the title.

## C Developmental Ratios

Number	Idiom	Ratio
13	‘Voor de hand liggen’	2.57
54	‘Lachen als een boer met kiespijn’	2.57
50	‘Hoe de vork in de steel zit’	2.00
7	‘Per slot van rekening’	1.71
3	‘In de steek laten’	1.20
24	‘Een brok in zijn keel voelen’	1.14
34	‘Een oogje in het zeil houden’	1.14
5	‘Op het punt staan om’	1.07
32	‘Er nog een schepje bovenop doen’	1.03
51	‘Te mooi om waar te zijn’	1.03
6	‘Van top tot teen’	1.00
8	‘Aan de slag gaan’	1.00
16	‘Een zucht van verlichting slaken’	1.00
20	‘Geen tijd te verliezen’	1.00
30	‘De waarheid tot hem doordringen’	1.00
1	‘Iets aan de hand zijn’	0.86
2	‘In de gaten hebben’	0.86
17	‘De stuipen op het lijf jagen’	0.86
38	‘Zijn woorden kracht bij zetten’	0.86
52	‘In geen velden of wegen te bekennen’	0.86
41	‘Van vlees en bloed’	0.80
10	‘Zijn hart klopt in zijn keel’	0.75
43	‘Het einde van de wereld’	0.71
4	‘In de rede vallen’	0.40

**Table C.1: Developmental ratios for each idiom, from high to low ratios. The number indicates the position of the idiom in the frequency list by Arbeek (2024), where low numbers correspond to a high position and frequency. Idiom numbers 25 and 21 are not included in the table, because of their infinite ratios.**