Article

Adding a Piece to the Puzzle: Children's Exposure to Idioms

Jacolien van Rij [1,*], **Floris H. Uithof** [2], **Sanne Poelstra** [3], **Stephen M. Jones** [1] and **Simone A. Sprenger** [4,*]

1   Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen, 9747 AG Groningen, The Netherlands
2   N.V. Nederlandse Gasunie, 9727 KC Groningen, The Netherlands
3   Ilionx, 9728 JR Groningen, The Netherlands
4   Center for Language and Cognition Groningen, University of Groningen, 9712 EK Groningen, The Netherlands
*   Correspondence: j.c.van.rij@rug.nl (J.v.R.); s.a.sprenger@rug.nl (S.S.)

Abstract: Idioms are figurative multiword expressions that need to be learned as part of the native phrasal vocabulary. While it has been shown that non-figurative multiword expressions are acquired with language exposure, the learning process for idioms may be different because the figurative meaning adds complexity to the learning task. Idiom vocabulary overall develops relatively late, but it is unknown to what extent children are exposed to idioms, and what kinds of idioms they encounter. Here, we investigated children's idiom exposure and its effect on the development of idiom vocabulary in three studies: we explore the frequency of a well-tested set of Dutch idioms in a corpus of child literature, test idiom familiarity in a controlled setting in primary school children, and compare those findings to a set of online familiarity ratings. We find that children's idiom exposure differs from adult idiom exposure, when comparing idiom frequencies based on children's books and a corpus with resources for adults. Idiom decomposability and idiom frequencies from the children's books, but not frequencies from the adult corpus, influenced the familiarity ratings of older children, suggesting that language exposure and idiom characteristics, such as decomposability, both play a role in idiom acquisition.

Keywords: idioms; language acquisition; language exposure; vocabulary; figurative language

1. Introduction

In addition to single words, a large part of our vocabulary consists of multiword expressions, such as *how are you* (greeting) or *know by heart* (to be able to retrieve some information from memory). These multiword expressions are estimated to comprise between 20% and more than 50% of our spoken and written language (e.g., Biber et al. 1999; Erman and Warren 2000). It has been suggested that these fixed phrases are easier to store in memory (e.g., Wray 2002) and have been found to facilitate production (e.g., Arnon and Cohen Priva 2013; Tremblay and Tucker 2011) and processing (e.g., Arnon and Snider 2010; Conklin and Schmitt 2012; Tremblay et al. 2011). Idioms are a subtype of these multiword expressions, as they are not only fixed phrases, but also have an intended meaning that is different from their literal meaning. For example, the meaning of *know by heart* is unrelated to the literal meaning of heart. Idioms are a fundamental part of our adult language and idiom vocabulary is crucial for nativelike language proficiency (Pawley and Syder 1983). For language learners—whether native or a foreign language (e.g., Cieślicka 2006; Conklin and Schmitt 2008)—this additional, idiomatic interpretation seems difficult to learn: the development of the idiom vocabulary seems to be delayed compared to the single word vocabulary (Carrol 2023; Kuiper et al. 2009; Sprenger et al. 2019). Whereas single-word learning levels off around the age of 20 (Brysbaert et al. 2016), idiom vocabulary levels off at a later age (Carrol 2023; Sprenger et al. 2019) or may even show a different development trajectory over age (Carrol 2023, but see Sprenger et al. 2019). It is an open question what

the mechanisms are that lead to the observed delay in idiom acquisition, relative to the single-word vocabulary. This paper aims to add more insight in the process of idiom acquisition by investigating children's exposure to idioms.

Unlike single-word vocabulary research, attempts to quantify idiom knowledge so far have been sparse. We know neither how many idioms an adult language speaker knows, nor how many idioms are part of the language. Idiom dictionaries may contain more than 10,000 idioms (e.g., Ayto 2010 for English and de Groot 1999 for Dutch), but it is not clear whether language users know and use all of these. Martinez and Schmitt (2012) constructed a list of 505 'phrasal expressions'—non-transparent and fixed multiword expressions with high frequencies—which may include syntactically fixed, non-transparent idioms. They concluded that these multiword expressions are part of the top 5000 most frequent word families. This led Brysbaert et al. (2016) to calculate that a 20-year-old English native speaker knows on average 4200 non-transparent multiword expressions and a 60-year-old 4820 (i.e., 10% of the estimated average single-word vocabularies of 20- and 60-year-olds). However, these estimates assume that the proportion of multiword expressions and idioms in our vocabularies is constant over age, which is not in line with the above-mentioned finding that idiom knowledge develops slowly across the life span (Carrol 2023; Sprenger et al. 2019). Furthermore, we also do not yet know much about children's and adolescents' idiom vocabulary.

The prevalence of multiword expressions in languages has inspired an ongoing debate on how language is learned and represented: various studies find evidence that children store frequently co-occurring words (i.e., multiword expressions) as holistic representations, and only later learn to decompose these (e.g., Arnon and Snider 2010; Bannard and Matthews 2008; see Arnon and Christiansen 2017, for a review). This contrasts with the traditional view that children first learn words and apply computational operations to combine these into larger units (e.g., Pinker 1991, see Contreras Kallens and Christiansen 2022, for a recent theoretical approach to this problem). For example, Bannard and Matthews (2008) reported that 2- and 3-year-olds are more accurate and faster in repeating high phrase frequency four-word expressions than similar four-word expressions with a low phrase frequency, using multiword expressions extracted from a corpus of child-directed speech. Similar processing advantages for multiword expressions have also been reported for adults (e.g., Arnon and Snider 2010; Arnon and Cohen Priva 2013; Tremblay et al. 2011). The finding that young children are sensitive to the phrase frequency of multiword expressions suggested that these phrases are stored in memory from a young age. Interestingly, a study by Nicoladis (2019) suggests that children can decompose fixed expressions at an earlier age than generally assumed: Nicoladis analyzed highly frequent fixed expressions produced by a three-year-old French-English bilingual child and found cross-linguistic influences on the child's use of fixed expressions (for example, using the expression 'I have hungry' instead of 'I am hungry' in English, because in French the verb *avoir*—'to have'—is being used for this expression).

As idioms are also multiword expressions, the above-mentioned findings suggest that *frequency* should be a good predictor for idiom knowledge in children. However, the figurative meaning of idioms adds another dimension to the task of the learner, because they do not only need to learn a specific configuration of words, but also their non-literal interpretation. It is conceivable that the difficulty of this task varies with the degree to which an idiom is perceived as being *transparent*: the more straightforward the relationship between the idiomatic meaning and the literal meaning of the constituent words, the more transparent the idiom. Various experimental studies have shown that transparent idioms are easier to understand and explain for children (e.g., Cain et al. 2009; Gibbs 1991; Levorato and Cacciari 1992; Nippold and Taylor 1995; Nippold and Rudzinski 1993). For opaque idioms, without a clear relation between the two different meanings, there is a consensus that language users must have stored and be able to retrieve the idiomatic meaning in memory in order to arrive at the correct interpretation (e.g., Swinney and Cutler 1979). In contrast, various studies have shown that language users—in the case of unfamiliar

transparent idioms, even children as young as eight years old (Levorato and Cacciari 1992)—can infer the idiomatic meaning, given a supporting context (e.g., Cain et al. 2009; Gibbs 1987, 1991; Levorato and Cacciari 1992; Nippold and Martin 1989). This suggests that the idiomatic meaning can also be constructed compositionally, based on the constituent words (e.g., Cacciari and Tabossi 1988; Titone and Connine 1999). Titone and Connine (1999) have argued that in idiom processing, both strategies—retrieving the idiomatic meaning from memory and interpreting the idiom compositionally—may be applied, and that the success of the strategies depends on the idiom characteristics: frequent and/or opaque idioms may benefit from storing the meaning in memory, whereas the compositional process may work better for the less common and/or more transparent idioms. Thus, both frequency of occurrence and the idiom's transparency may influence idiom processing and memory access.

These same two predictors—idiom frequency and transparency—also play an important role in theories of idiom acquisition. One line of research has focused on children's exposure to idioms, measured by familiarity ratings (e.g., Nippold and Taylor 1995; Nippold and Rudzinski 1993; Nippold and Martin 1989). In these experiments, children and adolescents show a gradual increase in their understanding of idioms with age, and better comprehension for high-familiarity idioms than for low-familiarity idioms. Interestingly, Reuterskiöld and Van Lancker Sidtis (2013) found that children do not require much exposure to learn unfamiliar idioms: 8–9-year-old children recognized idioms significantly better than novel utterances that they had heard once in a conversational context. Another line of research has focused on how children learn to derive an idiom's meaning from context. Levorato and Cacciari (1992, 1995) have proposed that the compositional interpretation of idioms is dependent on children's cognitive development: children first need to acquire *figurative competence*, a set of skills necessary for considering a figurative interpretation and for deriving a figurative interpretation from the context. Only with figurative competence are children able to make use of the idiom's transparency to compute the figurative meaning. Figurative competence develops over time, starting around age eight years old (Levorato and Cacciari 1995). This set of skills is not specific to idiom interpretation, but is assumed to facilitate language comprehension in general, and may include semantic analysis—i.e., retrieving alternative meanings of polysemous words and inferring a non-literal meaning of a phrase—and context processing—i.e., deriving an interpretation that is coherent with the broader context (e.g., Cain et al. 2009; Levorato and Cacciari 1995). Levorato and Cacciari (1995) proposed that children focus on the literal meaning of the words (in a 'local, piece-by-piece elaboration of the text', Levorato and Cacciari 1995), and need to learn to integrate information from the global discourse to arrive at an interpretation that is coherent with the surrounding context. Other researchers (e.g., Piquer-Píriz 2020; Winner 1988; Zurer-Pearson 1990) have argued that figurative competence is acquired gradually, starting at a much earlier age. It has been proposed that one of the reasons for young children's tendency to interpret figurative language literally may be that they lack world knowledge to understand the link between the figurative and literal interpretation (e.g., Piquer-Píriz 2020). An implication of this hypothesis is that children may understand and use a very specific set of idioms that fits their understanding of the world. Furthermore, for this specific set of idioms young children may be able to use transparency to derive the figurative meaning from context. This may explain why even 5-year-old children have been found to select the correct interpretation of idioms when presented in a supportive context (Cain et al. 2009; Gibbs 1987).

Although these lines of research are sometimes contrasted, it is probable that both idiom exposure and the development of figurative competence skills (which may be dependent on increasing world knowledge, cf. Piquer-Píriz 2020) jointly play an important role in idiom acquisition, and that their relative contributions may be modulated by idiom characteristics such as transparency (comparable with Titone and Connine (1999)'s proposal for idiom *processing*).

1.1. Current Study

In our view, idioms pose two challenges for learners of a native language: (1) language learners need to detect that a multiword expression is idiomatic, having another meaning than the literal (compositional) meaning of the constituent words, and (2) they need to learn the intended meaning. Figurative competence plays a role in these two processes: children need to be aware that phrases can have a non-literal meaning in order to detect an idiom, and they need sufficient context processing skills and world knowledge to derive the intended meaning from the context (cf. Cain et al. 2009; Levorato and Cacciari 1992, 1995; Piquer-Píriz 2020). Our hypothesis is that when children's figurative competence is developing (i.e., in older children), idiom transparency will start to influence their learning and processing: transparent idioms are understood better and remembered more easily, and therefore, they become relatively more familiar than opaque idioms. In addition, we expect an influence of idiom frequency for the same children, because language exposure also plays a role in these two processes (Nippold and Taylor 1995): children need to encounter an idiom in order to recognize it as such, and they need to learn the idiomatic meaning.

However, these two processes are only necessary when the idiom actually has a competing literal interpretation. When the idiom is only used idiomatically, children may just learn to associate this idiomatic meaning with this phrase, similar to learning other types of multiword expressions (cf. Arnon and Christiansen 2017; Arnon and Snider 2010; Bannard and Matthews 2008). In such learning situations, figurative competence (i.e., the ability to derive the figurative meaning) is not required and language exposure will be the only predicting factor. As a result, our hypothesis is that young children may show an effect of frequency, but not for transparency, and only for specific idioms that occur in children's language input.

To test these hypotheses, the current study investigates the effects of idiom exposure on children's acquisition of idioms in their native language, and also the influence of transparency as a marker of children's figurative competence.

1.2. Operationalizing Idiom Frequency and Transparency

Levorato and Cacciari (1992) have put forward a similar proposal and have tested the roles of idiom familiarity and context on idiom interpretation in seven-year-old and nine-year-old children. The seven-year-olds showed more idiomatic interpretations for familiar idioms, whereas familiarity did not play a large role for the nine-year-olds when context was provided, suggesting that younger children may be more sensitive to exposure than older children.

A potential problem with this study and other studies is that frequency of occurrence is operationalized in terms of familiarity ratings (e.g., Levorato and Cacciari 1992; Nippold and Taylor 1995, see also Bonin et al. 2013; Hubers et al. 2019; Tabossi et al. 2011; Titone and Connine 1994). Typically, these familiarity ratings are provided by adult participants (for example, by teachers in Levorato and Cacciari 1992). However, idiom familiarity increases with age and children's familiarity ratings are generally quite different from those of adults (Carrol 2023; Nippold and Rudzinski 1993; Sprenger et al. 2019). Therefore, adults' familiarity ratings may not reflect children's idiom exposure. An additional theoretical concern is that the use of familiarity ratings to capture frequency of occurrence assumes that idioms are always stored in memory. However, it may be easier to store an idiom for which the idiomatic interpretation was successfully derived from context than an opaque idiom that was not understood well (however, see Reuterskiöld and Van Lancker Sidtis 2013). It also has been found that familiarity may not solely capture frequency of occurrence, but may also be influenced by transparency and other idiom characteristics (cf. Carrol et al. 2018; Keysar and Bly 1995; Nordmann et al. 2014). For example, Carrol et al. (2018) showed that for native speakers, familiar idioms are perceived as more transparent. Therefore, the current paper investigates children's idiom exposure by looking at corpus frequencies (cf. Bannard and Matthews 2008, for multiword expressions).

To date, few studies have used corpus frequencies to capture idiom exposure. A probable reason is that idioms are not easy to find in a corpus, because they may show many types of variation (including syntactic and/or lexical variation, as well as insertions or modifications of adjectives and adverbs, e.g., Barlow 2000; Moon 1998). Furthermore, corpus studies require at least some degree of manual checking of the results, to verify that the non-literal meaning of the phrase was intended. Sprenger et al. (2019) collected frequency counts for 193 Dutch idioms from the Lassy Large corpus (Van Noord et al. 2013), a 700-million-word corpus of Dutch texts from mixed sources. They investigated how frequency and *decomposability* (i.e., a different measure to quantify the relation between the literal and figurative interpretation, rating how strongly the literal meaning of the constituent words contribute to the figurative interpretation) influenced familiarity ratings for different ages, and reported interactions between age and frequency and age and decomposability: whereas *low-frequency idioms* receive low familiarity ratings for all ages, *high-frequency idioms* show a sharp increase in familiarity ratings over age for participants younger than 30 and are rated as being highly familiar by participants older than 30 (Sprenger et al. 2019). Decomposability seemed to only influence familiarity ratings by young adults, with the familiarity ratings increasing with decomposability (i.e., higher decomposability ratings reflect a shorter distance between the literal and figurative interpretation). The current paper uses these same idioms with frequency and decomposability ratings to compare children's and adults exposure to these idioms.

In the literature, the relation between the literal meaning of the constituent words and the idiomatic interpretation has been defined and measured in different ways. For example, the concept of *transparency* measures how easy it is to derive the idiomatic meaning from the literal meaning, focusing on the underlying motivation (i.e., why this idiomatic meaning is associated with the phrase Cieślicka 2015). Another commonly used measure is *decomposability* (or semantic analyzability), which measures how strongly the literal meaning of the constituent words contribute to the figurative interpretation, focusing on the constituents and structure of the idiom (Cieślicka 2015). However, these measures are not defined in consistent ways and the terms are used interchangeably (as discussed in Carrol et al. 2018; Hubers et al. 2019). Typically, the idiomatic meaning is provided when these ratings are collected to control the idiomatic meaning that participants rate (e.g., Hubers et al. 2019; Sprenger et al. 2019; however, see Carrol et al. 2018, for another approach). Although transparency and decomposability measure different idiom properties (e.g., Carrol et al. 2018; Cieślicka 2015; Nunberg et al. 1994), they are related in that they both aim to quantify an aspect of the distance between the literal and figurative interpretation. For the current studies, we assume that when children's figurative competence is developing (i.e., in older children) they are able to use both idiom properties—transparency and decomposability—to derive the figurative meaning from context. Therefore, both measures can serve as a marker for the development of children's figurative competence. In this study, we have used the collected decomposability ratings from Sprenger et al. (2019) to quantify the relation between the literal and idiomatic meaning, and therefore we will use the term **decomposability** in the remainder of this paper, with higher values indicating a stronger relation—that is, with higher decomposability values, it is easier to get to the figurative meaning from the literal meaning.

In the following sections, we will present three studies that together investigate children's exposure and familiarity with idioms. Study 1 is a corpus study investigating which idioms of the database of (Sprenger et al. 2019) occur in a corpus of Dutch children's books. Study 2 is a controlled experiment that compares children's and adults' familiarity with idioms, and how these are influenced by frequency counts from the adult corpus, the frequency counts from the children's books corpus, and idiom decomposability. Study 3 presents familiarity ratings from children collected by means of an online questionnaire, using the same set of idioms, to verify the results of Study 2. Together, these three studies provide new insights in children's exposure to idioms and how their idiom exposure influences their idiom vocabulary as measured in the familiarity ratings.

2. Study 1: Corpus Study

The aim of this study is to investigate whether and how children's exposure to idioms is different from adults' exposure. To answer this question, we determined idiom frequencies in a corpus of Dutch children's books and compared them to those found in the Lassy Large corpus (Van Noord et al. 2013).

2.1. Methods

We started with the idiom database created by Sprenger et al. (2019), which consists of adult frequency measures and decomposability ratings for 193 Dutch idioms. The idioms in this database each have one or two nouns, but are not controlled for syntactic structure or position of the nouns. In the original study, all items were presented in past tense and preceded by the temporal adverb 'Toen' (*then*, at a time in the past), for example 'Toen hield hij een oogje in het zeil.' (*Then he held an eye in the sail*, which means *to keep an eye on things*.) Four idioms were translated from German as control items (see Sprenger et al. 2019). We included these idioms in our current study, because at least three of these have been found to occur in the Dutch language too, albeit with very low frequencies. For all idioms, Sprenger et al. (2019) obtained frequencies from the Lassy Large corpus (Van Noord et al. 2013), a 700-million-word corpus of Dutch texts with automatically assigned syntactic annotations that is composed of both spoken and written subcorpora. We will refer to these frequency counts with the term **adult frequency**. In addition, Sprenger et al. (2019) collected decomposability ratings for all idioms in an online questionnaire: participants were asked how strongly the constituent words contributed to the idiomatic meaning of the phrase. The decomposability scores were derived from the ratings of 34 native Dutch participants in the age range 21–26 years old (mean 24.3 years old). We excluded one idiom from the database, because there was a misspelling in the presentation of the idiom in the decomposability questionnaire.

To investigate children's exposure to the remaining 192 idioms from the database of Sprenger et al. (2019), we built a corpus of 50 Dutch children's books. The books were selected such that they cover different age groups, based on the book categories from the Dutch public libraries (0–6 years: AB (baby), AP (toddler), AK (preschool), 6–9 years: A, 9–12 years: B, 12–15 years: C, and 15–18 years: Young Adult). Our aim was to include popular and well-known children's books to form a representative corpus, and accordingly, we took advice from the local library and a primary school. Table 1 below lists the number of books for each target age group and the number of words in each collection of books. The books for younger children typically contain fewer words, and therefore, they contribute less to the corpus. Therefore, we decided to combine the first three target age groups (target age 0–12 years) for our analyses below, so that we can compare three collections with comparable size (i.e., books for 0–12 years, 12–15 years, and 15–18 years).

Table 1. Overview of the number of books for the different target age groups and the number of words in each collection of the children's books corpus.

Target Age	Books	Number of Words (%)		Analysis
0–6	6	4216	(0.16%)	
6–9	11	241,297	(9.41%)	**0–12:** 34.1%
9–12	11	628,888	(24.53%)	
12–15	11	797,543	(31.10%)	**12–15:** 31.1%
15–18	11	891,988	(34.79 %)	**15–18:** 34.8%
Total:	50	2,563,932	(99.99%)	

After the books were scanned, the tool SketchEngine (Kilgarriff et al. 2014) was used for Optical Character Recognition (OCR) and to search for the idioms using Contextual Query Language (CQL). Because of copyright reasons, we cannot make the corpus publicly

available, but a full list of book titles is provided in Appendix A.2. Most idioms were searched based on the combination of the lemmas of the nouns, to allow for variation in form, and afterwards, the results were manually checked. To reduce the number of results, sometimes, non-optional characteristics of the idioms were included in the CQL query.

The frequency counts for the different corpora (LassyLarge corpus, Van Noord et al. 2013, the children's books corpus), and for the different subcollections of the children's books corpus (i.e., books for the three age groups) were converted to a log scale using the Zipf-transformation (Van Heuven et al. 2014; cf. Carrol 2023).[1] The Zipf-transformation normalizes the frequency per million words to a range between 0 and 7, which facilitates comparing the frequencies from the large corpus of adult texts and the much smaller children's books corpus. We labeled the idioms that were not found in the corpora and collections with −1.

2.2. Analyses

We used Generalized Additive Mixed Modeling (Hasties and Tibshirani 1990; Lin and Zhang 1999; Wood 2017) as implemented in the R package 'mgcv' (version) 1.8.41 (GAMMs; Wood 2017) for all analyses in this paper, and the package 'itsadug' (version 2.1.4, van Rij et al. 2022) for interpretation and visualization of the results. GAMM is a non-linear mixed-effects regression technique that allows for a non-linear relation between the dependent variable and a predictor or a combination of predictors (interaction). The non-linear relation is determined by the data, and does not need to be specified a priori; when no non-linear relation is supported by the data, a linear trend is fitted. In addition, GAMM allows the inclusion of various types of random effects (random intercepts, random slopes, and non-linear random trends) to account for variability in participants and items. The three datasets discussed here were unbalanced, with limited observations per cluster, and therefore, non-linear random trends were not supported. When using GAMM, different methods are used to assess statistical significance: (1) summary statistics, (2) model comparison procedure comparing Maximum Likelihood scores and AIC scores, and (3) visualization and inspection of the full model (cf. van Rij et al. 2019; Wieling 2018). The reason for not relying on model comparisons alone is that model comparison procedures are not always reliable if models are not strictly nested.

To facilitate the readability of the text and to reduce the amount of statistical information in the paper, we summarize the results here and provide only the most essential statistical details. For the interested reader, the data and all analyses are provided as Supplementary Materials at https://git.lwp.rug.nl/p251653/childrens-idiom-exposure.

2.3. Results

Retrieved idioms. Figure 1a shows how many idioms were retrieved from each of the three collections in the children's books corpus and from the adult corpus. Interestingly, the number of idioms that was retrieved from the children's book corpus collections increases with target age: 40 different idioms were found in the books for 0–12-year-olds, 51 idioms were found in the books for 12–15-year-olds, and 64 different idioms were found in the books for the 15–18-year-olds. A three-sample test for equality of proportions without continuity correction suggests that the number of retrieved idioms in these three collections is significantly different ($\chi^2(2) = 7.64$, $p = 0.022$).

The sets of idioms retrieved from these three collections are different. Only 29 of the 40 idioms that occur in the books for children between 0 and 12 years old are also found in the books for older age groups (20 and 25 of the 40 idioms overlap with the idioms found in the books for 12–15 and for 15–18, respectively, see Figure 1a). The differences between these three collections of books are caused by the many idioms that are found only once, which comprise more than half of the retrieved idioms per collection (23, 28, and 38 idioms, see Figure 1b, which correspond to 57.5%, 54.9%, and 59.4% of the retrieved idioms, respectively). The many single observations—and as a consequence, the lack of overlap between the retrieved idioms—are due to the relatively small size of the collections

of the children's books corpus. When we collapse the three collections, the percentage of retrieved idioms increases to 49.4% (95 of 192 idioms were retrieved) and the number of idioms that was retrieved only once decreases to 38.9% (37 out of 95 retrieved idioms, see Figure 1b). Figure A1b in Appendix A shows the distribution of frequency counts: 21 idioms were retrieved two times, 31 idioms 3–8 times, and 6 idioms were retrieved 10 times or more (max 39 times).
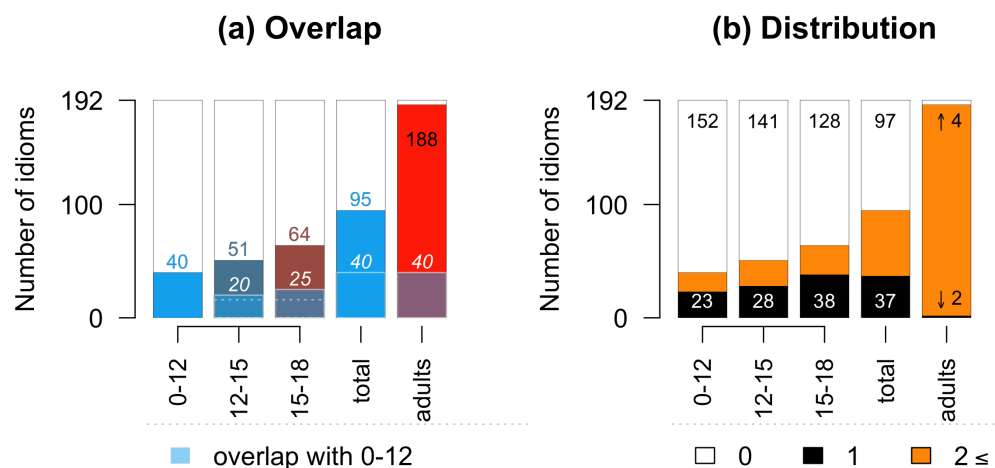
(a) Overlap   (b) Distribution

Figure 1. Idioms retrieved from the children's books corpus and from the Lassy Large corpus. Plot (**a**) shows the number of retrieved idioms for the three collections in the children's books (first three bars from the left), and the two whole corpora, children and adults (rightmost bars). The blue lower bars show the overlap with the retrieved idioms from the collection of books for the youngest children (age 0–12). Note that throughout the paper the blue color represents children and the red color represents adults, and increasing age is represented by the color gradient from blue (youngest children) to red (adults). Plot (**b**) shows the number of idioms that were not retrieved (white area at the top of the bars), the number of idioms that was retrieved once (black bars at the bottom), and the number of idioms that were retrieved more than once (orange parts in between the black and white).

Idiom frequency and decomposability. We investigated whether an idiom's decomposability was predictive for the adult frequencies and the frequency in the children's books corpus (henceforth **child frequency**). Only the non-zero frequency counts (i.e., Zipf frequencies larger than 0) were included in the analysis. We found that decomposability does not have a significant effect on the Zipf-transformed adult frequencies, but it does have a significant (linear) effect on the Zipf-transformed child frequencies ($F(1, 92) = 5.53$; $p = 0.021$), as illustrated in Figure 2a,b. However, the adult frequencies were not predictive for the child frequencies. These results suggest that children's and adults' idiom exposure may be quite different, with children being significantly more often exposed to transparent idioms.

As only 50.5% of the idioms in our sample occurred in the children's books corpus, we also tested whether decomposability and adult frequency predicted whether or not an idiom was found in the children's books corpus. The dependent variable is binary (0 = not found, 1 = found in children's books corpus), and accordingly, we used logistic regression (GAMM with family binomial) to test this. Both decomposability and adult frequency showed a significant effect: higher decomposability ratings increase the probability of an idiom appearing in the children's books corpus (non-linear trend, $\chi^2(1.82) = 7.66$, $p = 0.026$). The adult frequencies show a similar, but linear effect ($\chi^2(1) = 12.59$, $p < 0.001$). Figure 2c illustrates the two effects, transformed to proportion scale for interpretation. Note that the two predictors show considerable overlap in their predicted effect. To avoid spurious findings due to collinearity of these predictors, we analyzed the effects of the predictors separately, and these analyses converge to the same results. Taken together, the analyses

show that both decomposability and adult frequency contribute independently to the probability with which an idiom can be found in the children's corpus.
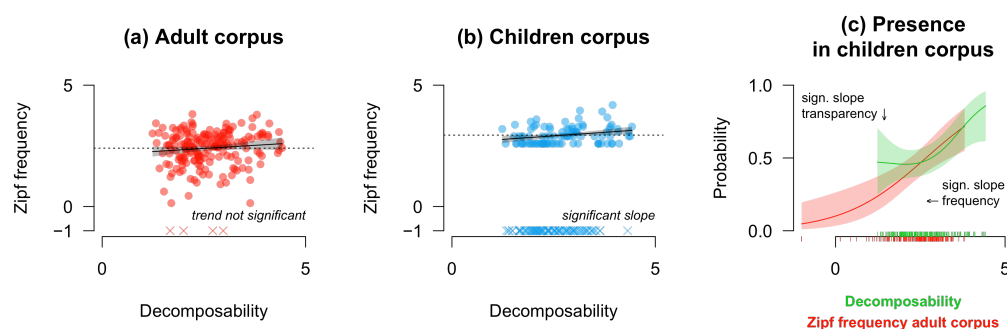

Figure 2. Results of analyses Study 1: (**a**) Relation between decomposability (on the x-axis) and adult frequency (Zipf-transformed; on the y-axis). The grey line is the estimated effect of decomposability. The red color refers to the frequencies of the Lassy Large corpus (cf. Figure 1). (**b**) Relation between decomposability (on the x-axis) and child frequency (Zipf-transformed; on the y-axis). The grey line is the estimated effect of decomposability. The blue color refers to the frequencies of the children's books (cf. Figure 1). (**c**) The effects of adult frequency (red line; x-axis) and decomposability (green line; x-axis) on the estimated presence of idioms in the children's books corpus (y-axis; binary dependent variable, modeled on logit scale; estimates are transformed to proportion scale). All model estimates are summed effects, with random effects excluded, with pointwise 95% confidence intervals.

Comparison of the three collections. We were interested to see whether the effects of decomposability and adult frequency are supported by all three collections in the children's books corpus. We expected that decomposability would most strongly predict the presence in books for older children (target ages 12–15 and 15–18), because children at these ages are expected to have the figurative competence skills required for using literal meanings to derive the intended meaning from the idioms (cf. Levorato and Cacciari 1992, 1995). We combined the data of the three collections to investigate whether the presence of the idioms in the three collections is differently affected by decomposability and adult frequency. There is no difference between the three collections in the effect of decomposability, although numerically, the slope for decomposability is steeper for the collection of children's books for the youngest age group (slope estimates: 0.92 for 0–12, 0.75 for 12–15, and 0.69 for 15–18, but these are not significantly different). In contrast, the adult frequency only seems to affect the presence of idioms in the books for older children (i.e., target ages 12–15 and 15–18), but not the presence of idioms in the collection children's books for the youngest children (see Figure A2b in Appendix A). However, the difference between age groups is not supported by a model comparison procedure. Based on this, we conclude that both decomposability and adult frequency have a significant and similar effect on the idiom being present in all three collections of children's books.

2.4. Discussion

Our corpus study using Dutch children's books suggests that the number of different idioms occurring in children's books increases with age. This finding cannot be explained by differences in corpus size, because the three collections of books with different target ages are comparable in size. We do not know whether the number of different idioms found in the children's books collection with the target age 15–18 is comparable to the number of idioms in an adult corpus, because the children's books corpus and the adult corpus (LassyLarge, Van Noord et al. 2013) are very different in size, and hence, these results cannot be compared directly.

Our second finding is that decomposability and adult frequency are both predictive for the probability of an idiom appearing at all in the children's books corpus, but that only decomposability—and not adult frequency—influences the frequency counts of the idioms

that actually occur in the children's books corpus. Our explanation is that when idioms are low in frequency in the adult corpus—which is a very large corpus!—then there is no reason to expect these idioms to occur in the children's books. Conversely, when the idioms are highly frequent in the adult corpus, the probability increases of these idioms appearing in the children's books corpus. This effect of adult frequency seems mostly to be seen in the collections of children's books for the older age groups.

The effect of decomposability is interesting: idioms with higher decomposability are more likely to occur in children's books, even in the collection of children's books with target ages 0–12, and idioms with higher decomposability also occur more often than idioms with a lower decomposability. This may indicate that the (adult) authors of children's books select idioms with a high decomposability, because they estimate that children will understand these more easily. An alternative explanation may be that idioms that are relevant for children's books relate to situations that are more concrete and less abstract, which may result in higher decomposability ratings.

To summarize, this corpus study suggests that children's exposure to idioms is quite different from that of adults, because the frequency values of the idioms they encounter are not influenced by frequency values from the adult corpus. Children are more likely to encounter decomposable idioms than adults. To investigate whether the observed difference in exposure is reflected in children's familiarity with idioms, we compared children's and adults idiom familiarity in a controlled experiment.

3. Study 2: Controlled Experiment

To test whether children's idiom vocabulary is predicted by children's idiom exposure and the idiom's decomposability, we asked children of around 7 years old, children of around 9 years old (cf. Levorato and Cacciari 1995), and adult controls to indicate their familiarity with 104 idioms from Study 1 (which were based on the Sprenger et al. (2019) database). For these items, decomposability ratings and adult frequency counts are available, and additionally we added the frequency counts from Study 1 (henceforth **child frequency**). Children that were between 7 and 9 years old were tested, because we expect to see differences between these age groups. Experimental studies suggest that 7-year-old children generally have more difficulties with selecting the correct interpretation for idioms *without supporting context* than 9-year-old children (e.g., Cain et al. 2009; Gibbs 1991). We are interested to test whether this differences also shows in their familiarity ratings and whether the idiom frequency modulates their familiarity ratings differently. Furthermore, it has been proposed that figurative competence develops between 7 and 11 years old (cf. Levorato and Cacciari 1995). Therefore, we want to test whether 7- and 9-year-old children show a different effect of decomposability on their familiarity ratings.

3.1. Methods
3.1.1. Participants

Two classes from a Dutch primary school participated in the study: 15 participants (9 male, 6 female) from grade 4 and 15 participants (11 male, 4 female) from grade 6 in the Dutch school system, which means that the children were about seven years (m = 7;4 years, range = 6;11–7;9) and about nine years old (m = 9;7, range = 9;2–10;9), respectively. In addition, 15 adults (9 male, 6 female) participated as controls, with a mean age of 23 (range = 19–27).

3.1.2. Experimental Design

Each participant was presented with a unique semi-random list of 30 items: the idioms were ordered by frequency (based on the LassyLarge (adult) corpus; Van Noord et al. 2013) and labeled as *high frequency* (idioms 1–16, 15% of), *mid frequency* (idioms 17–47, 15–45%), and *low frequency* (idioms 48–100, 45–96%). For each participant, 15 idioms were randomly selected from the high-frequency idioms, 8 idioms were selected from the mid-frequency idioms, and 3 idioms were randomly selected from the low-frequency idioms. The reason

for using this semi-random selection procedure was to make it more likely for children to hear a familiar idiom—which would make the test more motivating—by including more higher frequency idioms. Due to an implementation error, the randomly sampled high-frequency idioms were presented first, followed by a block of randomly sampled mid-frequency idioms, and the experimental session concluded with a block of low-frequency idioms. As the experiment was rather short and the idioms were randomly ordered within these blocks, this unintentional effect probably did not have large consequences on the results. Each list also contained four control items in fixed positions in the list, namely at trials 7, 14, 21, and 28. These control items were idioms that were not expected to be familiar to children, because they had low frequencies in the adult corpus and they were not familiar to the (adult) Dutch speaking authors involved in designing this study.

3.1.3. Procedure

The experiment was implemented in Open Sesame 3.2.5 (Mathôt et al. 2012) and was presented on a Lenovo 10″ TB-X103F tablet. Each trial started with a fixation dot, followed by a and an idiom phrase, such as 'Toen hield hij een oogje in het zeil' (literal translation *Then he held an eye in the sail*, meaning to keep an eye on things). Each experimental session started with three practice trials, which were not included in the analysis data. The test block consisted of 30 trials. On each trial, the phrase was presented on the screen and after 500 ms the phrase was also presented auditorily. The sound files were prerecorded by a native female speaker of Dutch. After hearing the sentence, the question 'Ken je deze?' (*Do you know this one?*) was added to the screen with a big green checkmark symbol (✓) on the right and a big red cross symbol (×) on the left, for answering 'yes' and 'no', respectively. Participants had to press these pictures to indicate whether they recognised the idiom or not. Before answering the question, participants could ask the experimenter to play the sound recording again. After rating the idiom for familiarity, a second question was asked about the idiom ('Who is likely to use this idiom?'), aiming to identify with which age groups children associate the idiom. As this question was difficult for children to answer and the data are outside the scope of this paper, we will only present the results of the familiarity ratings here.

3.2. Results

Participants completed between 14 and 30 items (mean 28.9; 46 responses were missing in total). In addition, five participants were presented with one duplicate trial because of a technical error. The responses for the second encounter have been removed, only the first encounter with the idiom was included in the data. This increased the missing responses to 51 (3.7% of 1350), resulting in valid data for 93 (out of 104) idioms. Figure A3 in Appendix B shows the number of observations per idiom. We then selected only those idioms that received more than two observations (19 idioms were excluded and 30 out of 1299 observations, 2.3% of the data), resulting in valid data for 74 idioms: 16 high-frequency idioms, 29 medium-frequency idioms, and 29 low-frequency idioms. In addition, we used the four control items to check whether participants were actually doing the task: it would have been very unlikely that children are familiar with all four low-frequency idioms, so we decided to exclude children when they indicated familiarity with all four control idioms. Children indicated that they were familiar with 0–2 of the control idioms, and adults between 0 and 3 of the control idioms (Grade 4 mean 0.67, Grade 6 mean: 0.73, Adults mean: 0.67). No participants were excluded, and the control items were included in the analysis data. Figure 3 shows the average familiarity ratings for each age group.
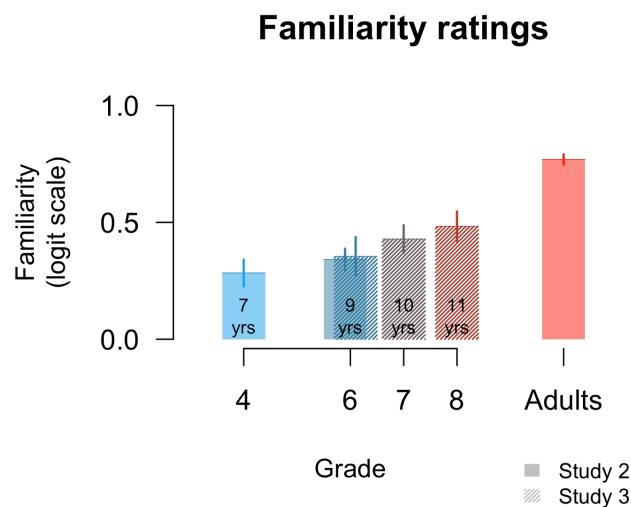
Figure 3. Average familiarity by school grade for Study 2 (solid bars) and Study 3 (dashed bars), with error bars indicating ± 1 standard error of the participant means. Study 2 tested children in grades 4 (7 years; light blue) and 6 (9 years; dark blue) and adult control participants (red). Study 3 includes children from grades 6 (9 years; dark blue), 7 (10 years; purple), and 8 (11 years; dark red).

To investigate the effects of decomposability, adult frequency, and child frequency on children's and adults' familiarity ratings, we ran two analyses. First, we ran separate analyses for each of the predictors, to see the individual contribution of each of these predictors and how it interacted with age group. We ran these analyses separately to avoid spurious effects due to collinearity of the predictors. For the second analysis, we reorganized the three predictors in orthogonal terms using principled component analysis (PCA), and included the three PCA components and their interaction with age groups in one model, to verify the results of our earlier analyses. In all analyses, children were grouped by their school grades instead of their age, because we did not have access to background information (such as IQ, verbal skills, and language or attention disorders). Children within a school grade may still show a large variation in language skills, but their attending a regular school program in the Netherlands ensures a minimal level of IQ and language experience. Figure 4 visualizes the results of the three separate analyses. Random intercepts for idioms and participants were included in all models to account for item and participant variability, but the data did not allow us to include random slopes for the predictors.

Decomposability. The top row of Figure 4 shows the effect of idiom decomposability on the familiarity ratings. The adult participants (right panel) show a significant linear trend for decomposability ($\chi^2(1) = 23.50$; $p < 0.001$), quite similar in direction to the effect of idiom frequency in the adult corpus. The trend for the children in Grade 4 (left panel) was not significant ($\chi^2(1) = 0.01$), but the trend for the children in Grade 6 (center panel) was significantly different from zero ($\chi^2(1) = 5.13$, $p = 0.024$). A model comparison procedure indicated that the interaction between decomposability and age group contributed significantly to the model ($\chi^2(4) = 24.25$, $p < 0.001$, ΔAIC = 25.9). Put differently, we see in all but the youngest age group that idioms are more likely to be familiar if they are also considered to be more transparent.

Frequency in adult corpus. The middle row of Figure 4 shows the effect of the Zipf-transformed adult frequencies on the familiarity ratings. The adult participants (right) show a significant linear trend for frequency ($\chi^2(1)=41.68$; $p < 0.001$), but the trends for the children in Grade 4 (left) and Grade 6 (center) were not significant ($\chi^2(1) = 0.41$, $p = 0.52$, and $\chi^2(1) = 3.30$, $p = 0.07$, respectively). A model comparison procedure indicated that the interaction between frequency and age group was significantly contributing to the model ($\chi^2(4) = 44.8$, $p < 0.001$, ΔAIC = 45.5). That is, the more frequent an idiom in the adult

corpus, the more familiar it is to the adult raters. However, this relationship is not seen in the two groups of children.

Frequency in children's books. The bottom row of Figure 4 shows the effect of the child frequencies on the familiarity ratings. The adult participants (right) show a significant linear trend for child frequency $\chi^2(1) = 8.79$; $p < 0.003$), quite similar in direction to the effects of the adult frequency and decomposability. Again, the trend for the children in Grade 4 (left) was not significant ($\chi^2(1) = 0.095$), but the trend for the children in Grade 6 (center) was significantly different from zero ($\chi^2(1) = 11.85$, $p < 0.001$). A model comparison procedure indicated that the interaction between frequency and age group was significantly contributing to the model ($\chi^2(4) = 18.06$, $p < 0.001$, $\Delta \text{AIC} = 19.5$). Note that for this analysis, we only included the idioms that appeared at least once in the children's book corpus (35 out of the 74 idioms).
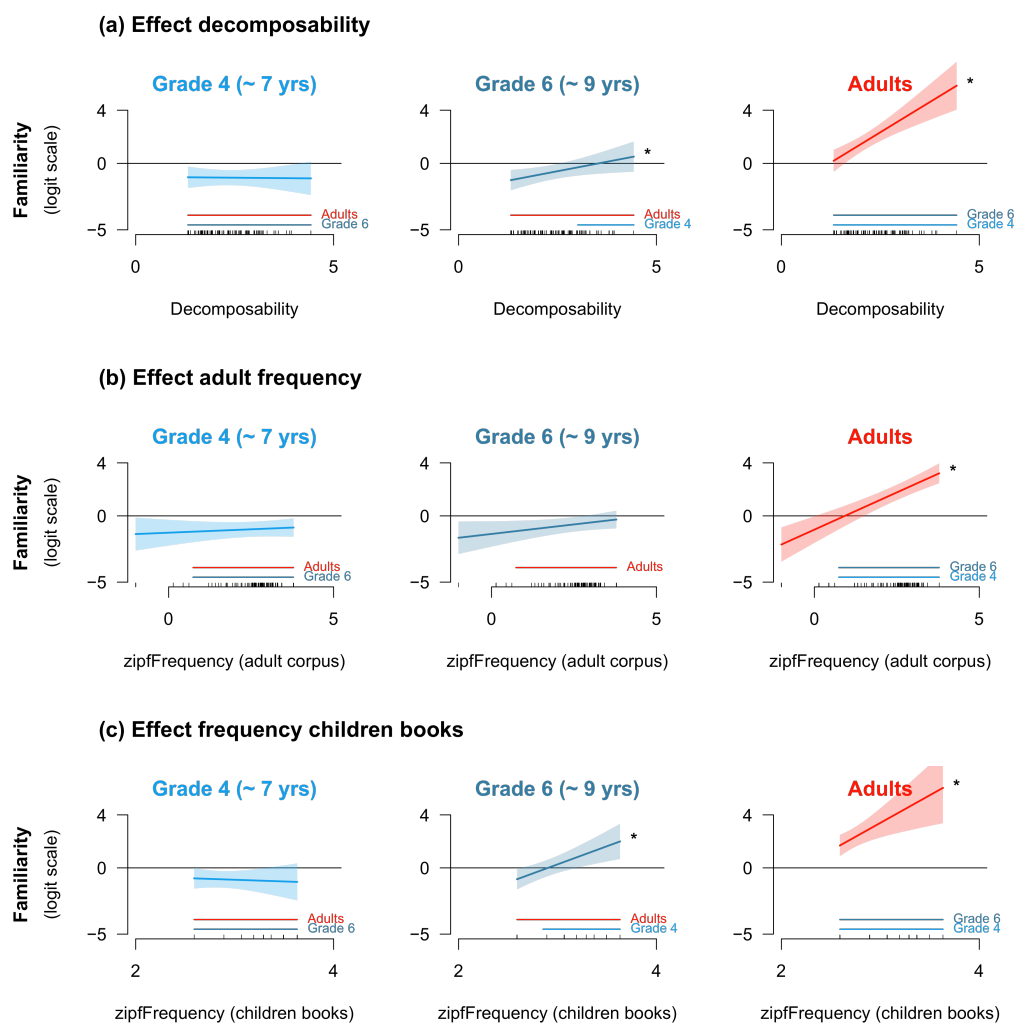

Figure 4. Estimated effects of (**a**) decomposability, (**b**) adult frequency (Zipf-transformed), and (**c**) child frequency on the familiarity ratings (y-axes) of Grade 4 children (left panels), Grade 6 children (center panels), and adult participants (right panels). The estimates are summed effects, with random effects excluded, with pointwise 95% confidence intervals. These estimates, differences, and confidence intervals are generated from three separate models, one for each predictor, with binary predictors modeling the (potentially non-linear) differences between the age groups. The solid horizontal lines close to the x-axes indicate significant differences with the indicated other age group. Significant effects (i.e., trends that are significantly different from a horizontal straight line) are marked with the symbol *.

Presence in children's books. We also tested whether the presence of an idiom in the children's books corpus (categorical predictor: 'yes', 'no') influenced the familiarity ratings, as a complementary measure of looking at the influence of children's idiom exposure. In this analysis, the familiarity ratings for the absent idioms are also included (74 idioms in total). This measure is illustrated in Figure 5: For the youngest children, there was no significant difference in their ratings for idioms that were present and absent in the children's books corpus. However, the ratings of the older children and adults increased significantly for idioms that were present in the corpus ($\beta_{\text{Grade6:PresentY}}$ = 1.051, SE = 0.354, z-value = 2.97, p = 0.003; $\beta_{\text{Adults:PresentY}}$ = 1.131, SE = 0.394, z-value = 2.87, p = 0.004).
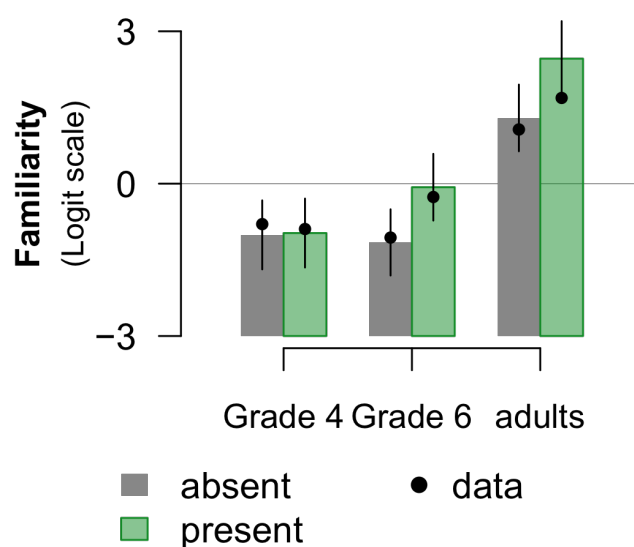

Figure 5. Estimated (summed) effects of the presence of idioms in the children's books (absence represented by gray bars, presence represented by green bars) on children's and adults' familiarity ratings, with pointwise 95% confidence intervals. The dots indicate the grand averages per condition, transformed to logit scale. Note that the model estimates are different, because these account for the unbalanced structure of the data by including random intercepts for items and participants.

PCA analysis. The analyses presented in Figure 4 show that all three predictors show a very similar influence on adult participants' familiarity ratings, and decomposability and child frequency seem to show the same effect for the older children (Grade 6). To test whether we can separate the effects of adult frequency, children's frequency, and decomposability, we reorganized the three (scaled and centered) predictors into three PCA components. All three components explain considerable proportions of the variance (0.47, 0.28, and 0.24, respectively), showing that they each potentially can account for variation in the data. The analysis only included the 35 idioms that were present in the children's books.

The model showed a gradual effect for PC1, which captures the shared effects of the predictors decomposability, adult frequency, and child frequency. Children in Grade 4 did not show a significant trend for PC1, but the familiarity ratings of children in Grade 6 and adult participants increased with increasing values for PC1. PC2 did not show a significant trend for any of the age groups and did not contribute to the model. Only the older children (Grade 6) showed a significant trend for PC3 ($\chi^2(1.0)$ = 6.21; p = 0.013). This component captures the difference between the child frequency and decomposability. The direction of the effect of PC3 indicates that the older children are more sensitive to child frequency than to decomposability.

3.3. Discussion

In contrast to our hypothesis, we did not find an effect of child frequency on the familiarity ratings of the youngest children. A potential reason may be that we did not include enough idioms that they knew, because the idiom list was originally constructed for adult participants (see Sprenger et al. 2019) and the selection procedure in this experimental study was based on adult frequencies. A closer look at the items that were rated by more than three children as familiar reveals that only one item fulfills this criterion for the youngest children, namely 'Toen hield hij een oogje in het zeil' (*Then he kept an eye on the situation*). For the older children, there are six idioms that meet this criterion. These idioms are listed in Tables A7 and A8 in Appendix B (with translations).

In line with our predictions, we found that older children's familiarity ratings are influenced by child frequencies, but not by adult frequencies. This confirms the conclusion from Study 1 that children's idiom exposure may be quite different from adult's idiom exposure. In addition, older children also showed an effect of decomposability, but the PCA analysis suggests that the effect of child frequency is stronger and may cancel out the effect of decomposability when these effects conflict.

Adults, on the other hand, showed an overlapping effect of adult frequencies, child frequencies, and decomposability. It may be the case that they are sensitive to all these three effects, or that these effects are driven by items for which these three predictors overlap in direction. Because these adult participants were relatively young, the results are in line with the results of Sprenger et al. (2019), who reported an effect of decomposability for younger adults.

In this study, the participants performed the task in the presence of the experimenter. This may have resulted in a response bias to rate idioms as familiar, even though the participants were explicitly instructed to only indicate idioms they recognized as familiar. However, the effects of frequency and decomposability are not expected to be cause or influenced by a response bias, because the participants were not aware of these manipulations. Nevertheless, it is useful to compare the overall ratings of Study 2 with Study 3, in which children completed an online questionnaire at home and did not meet the experimenter.

In Study 3, we zoom in more closely on the older children, comparing 9-, 10-, and 11-year-old children. We were interested to see whether the effects of decomposability and frequency would get stronger, and the children's familiarity ratings would become more adult-like with age.

4. Study 3: Online Questionnaire

To verify the results of Study 2, we compared the familiarity ratings of 9–12-year-old children in an online questionnaire involving 65 items from Study 1 (which were based on the Sprenger et al. (2019) data base). For these items, decomposability ratings and adult frequency counts are available, as well as the frequency counts from Study 1.

4.1. Methods

As part of the citizen science project *Maak dat de kat wijs*[2] (Sprenger and van Rij 2019), we invited several primary schools to participate in an online questionnaire to test the familiarity of Dutch idioms.

4.1.1. Participants

The questionnaire was available online. We provided primary schools with educational materials on idioms, and participation in the questionnaire was one of the assignments. Participants younger than 16 years old were asked for consent from parents or caregivers. Teachers and participants did not receive a reward for their participation.

4.1.2. Procedure

The questionnaire was available for computers, tablets, and smartphones. Participants had to complete a series of background questions first (including age, places of residence

and birth, and native language(s)) before starting with the idiom familiarity survey. Participants were instructed to indicate that they recognized the phrase by clicking on the 'JA' (*yes* in Dutch) button on the screen or the J on the keyboard, and the 'NEE' (*no* in Dutch) button on the screen or the F on the keyboard. They were also informed that some phrases were made up idioms. Children were presented with 30 phrases (including 4 fake idioms). The phrases were randomly selected out of a database with more than 1500 idioms and fake idioms that the first and last author created for the citizen science project.

4.1.3. Data Cleaning

For the current study, we selected all native Dutch primary school children who were also living in the Netherlands, and whose parents/caregivers had given consent to their participation in the study. This resulted in 134 participants. We only selected data from children attending the regular primary school system, to ensure that children did not suffer from severe language and attention disorders. We excluded four children who gave the same response to all trials (all unfamiliar: N = 3; all familiar: N = 1), because they may not have understood the task. We excluded two children, because they were the only children in grades 4 and 5 of the Dutch school system. The resulting 128 participants were 9–12 years old, and in grades 6, 7, and 8. The participants each completed 30 idiom ratings, including 4 non-existing (fake) idioms. The idioms were randomly selected out of a large variety of idioms, so that this group of participants in total rated 1099 idioms. The non-existing idioms were sampled from a list of 336 non-existing idioms. Participants were allowed to do the questionnaire multiple times, but only the first occurrence of each encountered idiom was included in the data.

For the current analysis, we only selected the idioms that were included in the Sprenger et al. (2019) study, because for these idioms decomposability ratings and various frequency measures were available. In total, 148 idioms from Sprenger et al. (2019) were included in the online questionnaire. By selecting only these idioms, the number of participants was reduced to 124, with each participant contributing 1–22 ratings (mean 3.4 ratings). In a next step, we only included idioms with three or more observations. This further reduced the number of participants to 116 (see Table 2), with each contributing 1–15 observations (mean 2.6), and 65 idioms, with 3–13 observations (mean 4.6).

Table 2. Participants per age group.

School Grade (Dutch School System)	Mean Age[3]	Age Range	N
grade 6 primary school	9.2	9–10	21
grade 7 primary school	10.4	9–11	50
grade 8 primary school	11.2	10–12	45

4.2. Results

Figure 3 shows the average familiarity ratings for each age group. The familiarity ratings are in the same range as those provided by the children from grade 6 (Dutch school system) in Study 2. We did not find significant differences between the three age groups.

We investigated the effects of decomposability, adult frequency, and the frequency in children's books in the same way as in Study 2. Figure 6 visualizes the estimates of the separate models for each predictor (which included all age groups combined in one model to allow for the interactions between the predictors and age groups). Random intercepts were included for items and participants, but random slopes were not supported by the data.
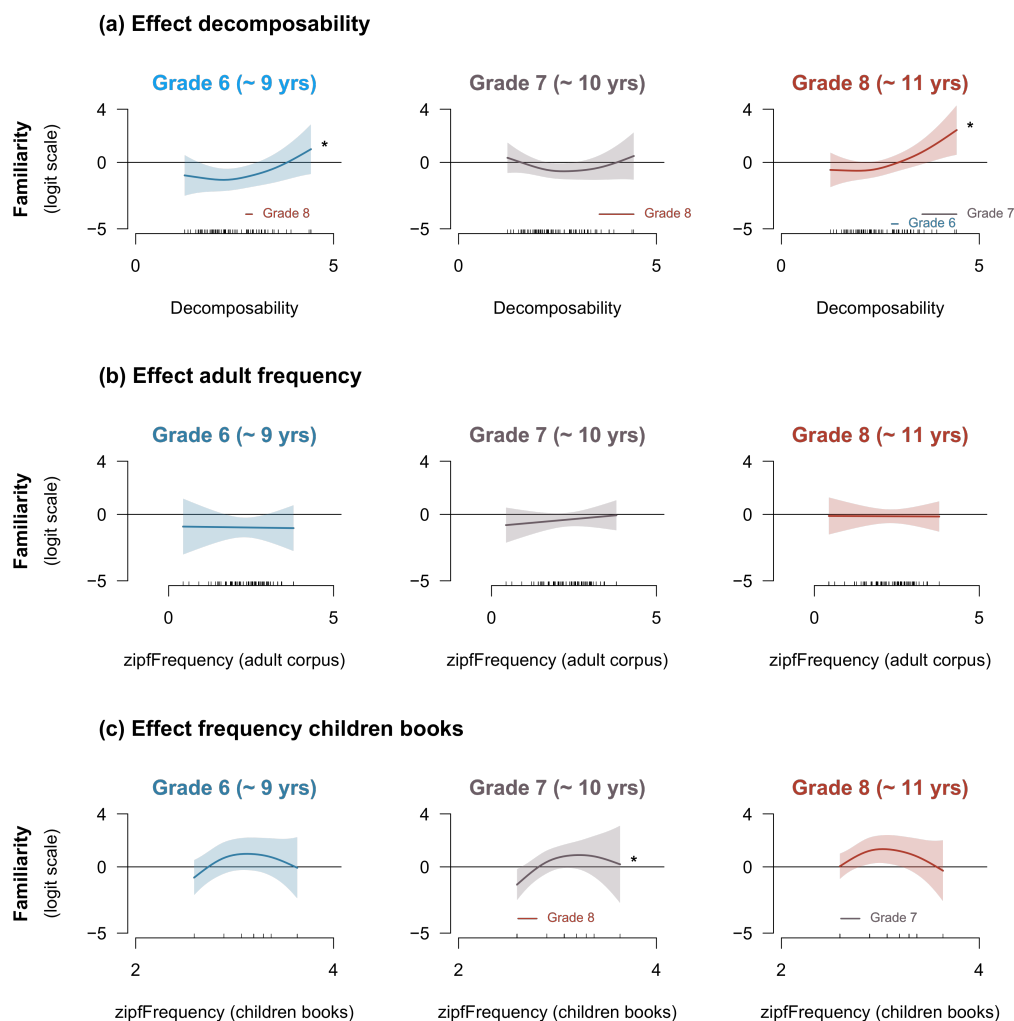
Figure 6. Estimated effects of (**a**) decomposability, (**b**) adult frequency (Zipf-transformed), and (**c**) child frequency on the familiarity ratings (y-axes) of Grade 6 children (left panels), Grade 7 children (center panels), and Grade 8 children (right panels). The estimates are summed effects, with random effects excluded, with pointwise 95% confidence intervals. These estimates, differences, and confidence intervals are generated from three separate models, one for each predictor, with binary predictors modeling the (potentially non-linear) differences between the age groups. The solid horizontal lines close to the x-axes indicate significant differences with the indicated other age group. Significant effects (i.e., trends that are significantly different from a horizontal straight line) are marked with the symbol *.

Decomposability. The top row of Figure 6 shows the effect of idiom decomposability on the familiarity ratings. The best-fitting model included a general non-linear trend for decomposability, which was significantly different from zero (summary statistics: $\chi^2(2.170) = 7.52$; $p = 0.036$). However, we did not find a difference in the trends for decomposability between the three age groups.

Adult frequency. The middle row of Figure 6 shows the effect of adult frequency on the familiarity ratings. We did not find any significant effect of adult frequency on the children's familiarity ratings.

Child frequency. The bottom row of Figure 6 shows the effect of idiom frequency in the children's books on the familiarity ratings (28 items). We did not find a difference in trend for child frequency between the three age groups. The best-fitting model included a general non-linear trend for child frequency, which was significantly different from zero ($\chi^2(2.168) = 11.85$; $p < 0.01$).

Presence in children's books. To verify that the effect of child frequency is not driven by the few items that were found in the children's books corpus and that have sufficient observations in the current dataset, we tested whether the presence in the children's books corpus ('yes' or 'no', categorical predictor) had an effect on the children's familiarity ratings (allowing us to include all 65 items). Figure 7 shows effects of the presence of idioms in the children's books on children's familiarity ratings, with the gray (absent) and green (present) bars representing the estimates from the statistical model, and the solid dots the grand averages of the data. This analysis again did not show a difference between the three age groups, but a main effect of an idiom's presence in the corpus (model comparison: $\chi^2(1) = 3.00$, $p = 0.014$): When an idiom was found in the children's books corpus, participants were more likely to rate this idiom as familiar ($\beta_{\text{PresentYes}} = 1.03$, SE = 0.32, z-value = 3.23, $p = 0.001$).
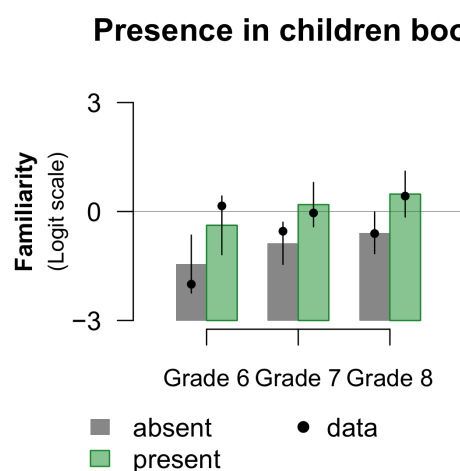
Presence in children books

Figure 7. Estimated (summed) effects of the presence of idioms in the children's books (absence represented by gray bars, presence represented by green bars) on children's familiarity ratings, with 95% confidence intervals around the estimates. The dots indicate the grand averages per condition, transformed to logit scale. Note that the model estimates are different from the grand averages, because these account for the unbalanced structure of the data by including random intercepts for items and participants.

Although fake idioms were excluded from the main analyses, we compared children's familiarity ratings on the fake idioms with their ratings of the idioms that were absent in the children's books and with their ratings of the idioms that were present in the children's books. The purpose of this comparison is to test whether fake idioms were evaluated differently from real idioms or not. The analysis based on participant familiarity counts for all 65 idioms suggests that both the idioms that were not found in the children's book corpus and the idioms that were present in the children's book corpus were rated significantly higher than the fake idioms ($\beta_{\text{PresentNo}} = 0.57$, SE = 0.22, z-value = 2.61, $p = 0.0091$; $\beta_{\text{PresentYes}} = 1.55$, SE = 0.240, z-value = 6.46, $p < 0.001$). Figure 8 illustrates the estimated familiarity ratings for the fake idioms in comparison with the existing idioms.

PCA analysis. Even though the analyses presented in Figure 6 do not show highly similar effects for the three predictors in the participant's familiarity ratings, we ran an additional PCA analysis to verify the results in a combined analysis. We reorganized the three predictors, after scaling and centering, into three PCA components, as in Study 2. Note that here, we again only used the 28 idioms that were present in the children's books. Only the first component, PC1, shows a significant effect on children's familiarity ratings (summary statistics: $\chi^2(1.0) = 3.48$; $p = 0.062$; model comparison: $\chi^2(2) = 5.228$, $p = 0.005$), but there was no difference between the three age groups. PC1 captures the shared effect of the three predictors adult frequency, child frequency and decomposability, and is supported by idioms for which these measures go in the same directions (for example, idioms showing

low values for all three measures). The positive linear trend of PC1 suggests that the effects of decomposability and frequency in children's books may be the same effect.
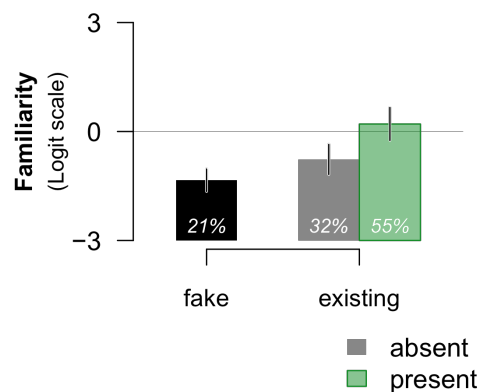
Comparison fake idioms

Figure 8. Estimated (summed) effects of the fake idioms (represented by the black bar) and the existing idioms (idioms not found in the children's books are represented by gray bars, idioms present in the children's books represented by green bars) on children's familiarity ratings, with 95% confidence intervals around the estimates on the logit scale. The labels on the bars indicate the estimated percentages of idioms that were evaluated as familiar.

4.3. Discussion

In Study 3, we investigated binary familiarity ratings for a subset of our idiom database. These data provide an additional, independent set of idiom familiarity ratings. A similar picture to Study 2 emerges: the familiarity of idioms in older children (9 to 12 years old) is influenced by decomposability and frequency in the children's books corpus, but not by adult frequency. Overall, however, the statistical effects of decomposability and frequency that we find are much weaker and less conclusive than the effects in Study 2. The reason is that the set of familiarity ratings that we used contained data for only a relatively small subset of our items (65 out of 192), with again relatively few observations per participant (1–15) and per item (3–13). As a consequence, the variation due to participants and items in this subset is more difficult to account for with statistical modeling. Nevertheless, we think that it is noteworthy that the pattern that emerges is similar to that found in Study 2.

In addition, the comparison between the fake idioms and existing idioms shows that the frequency effect exceeds potential response biases. The familiarity ratings of fake idioms (grand average of 23.5% familiar) indicate that children did not reject all fake idioms as unfamiliar, which suggests that children were at least somewhat more biased to rate idioms as being familiar than unfamiliar. Although the survey instructions explicitly explained that we were interested in which idioms children know and do not know, it is not unlikely that the familiarity ratings may have been influenced by a social desirability bias, for example to come across as knowledgeable. At the same time, the response biases are unlikely to cause the effects of child frequency and decomposability in Study 2 and Study 3, because the participants in these studies were not aware of these differences in frequency and decomposability. If anything, the response biases have made it difficult to detect the effect of idiom characteristics on their familiarity.

5. General Discussion

In the present work, we have investigated the extent to which idiom frequency and decomposability explain idiom knowledge in children between 7 and 12 years old. To this end, we combined adult frequency data and decomposability ratings from a previous study (Sprenger et al. 2019) with the results of three new studies: in Study 1, we determined the frequency of 192 Dutch idioms in a corpus of 50 popular children's books

(>2.5 million words) in order to determine the extent and quality of idiom exposure in children's literature. The results of this top-down approach show that only a subset of our items (i.e., less than half of them) appears in the corpus, often with very low frequencies. The sparseness of the data are in line with our expectations, as our item set was originally created for research on adults. Interestingly, however, we also see that the number of idioms that could successfully be retrieved from the children's book corpus increases with target age. These observations suggest two things. First, children are indeed exposed to idioms in children's literature, from the earliest ages onwards. Most probably, our estimates form a lower boundary for idiom exposure, as writers may very well have chosen to use other idioms that are not part of our item set. As a follow-up, it would be interesting to investigate idiom use in our corpus by means of a bottom-up approach, to see how many and what type of idioms are used in the corpus beyond our sample. Second, we see that the extent to which writers adapt their language use to their audience comprises the use of figurative language, with idiom use seemingly becoming more adult-like with target age. As writers—and, maybe even more so, editors—are strongly aware of their target audience, their use of idioms suggests that they expect their readers to be able to understand this type of figurative language, and that this understanding develops with age. Their expectations are in line with findings in the literature showing that children learn literal multiword expressions from a young age (e.g., Bannard and Matthews 2008), but also add the figurative dimension. To our knowledge, however, idiom knowledge in young children has not yet been studied systematically.

Another interesting conclusion from Study 1 is that frequencies of idiom occurrence in the children's books did not correlate with occurrence frequencies in the adult corpus. One consideration here is that the adult corpus is a corpus consisting of mixed sources, including fiction, spoken language, newspapers, Wikipedia entries, manual descriptions, and the annual speeches of the former Dutch queen, whereas the children's books corpus only consists of fiction texts, albeit written by 50 different authors. We nevertheless think that the difference between adult frequencies and the frequencies from the children's books reflects a difference in language exposure that we would have found when including other sources of children's language input, such as television programs and educational texts. Adults may also apply idioms to specific situations in which children are not involved: certain idioms may be commonly used in politics, but others in newspaper headlines or in business environments, all environments that are not part of children's contexts. For example, the Dutch idiom 'iets/iemand in de arm nemen' (literal translation *to take something/someone in the arm*, meaning to recruit someone or a company) is an idiom that is typically used in the context of hiring lawyers, detectives, construction companies, or gardeners. This idiom has a relatively high frequency in the adult corpus ($n = 2027$) but was not found in the children's books.

Apart from idiom frequency, Study 1 also investigated idiom decomposability. We found that the frequency in the children's books corpus increases with the idiom's decomposability. Put differently, children's authors seem to have a strong preference for decomposable idioms. This effect may be explained by the (adult) authors deliberately selecting idioms that are easier to interpret for children, who are developing the skills for interpreting figurative language. An alternative explanation could be that idioms that are relevant for children typically describe more concrete situations and require less specific world knowledge, and that these idioms are more decomposable, or are perceived as more transparent. In contrast to the children's book data, we did not find a similar effect of decomposability on the adult frequency counts. Interestingly, this finding mirrors observations by Carrol (2023), who collected transparency ratings from adults aged 18–77 years. Note that their transparency ratings are equivalent to our decomposability ratings. The transparency ratings were influenced by idiom frequency counts: the more frequent an idiom, the higher the transparency ratings, irrelevant of age. In other words, idioms that were rated as more transparent, were found more frequently in the corpus. The frequency counts were retrieved from a corpus of recent web-based newspapers and

magazines aimed at adult readers, Corpus of News on the Web (NOW; Davies 2016). The two corpora—the children's books corpus and the NOW corpus—have in common that the texts involve careful editing. Therefore, it seems a likely explanation that language users actually use more transparent idioms when carefully writing and editing their text. The alternative hypothesis, that language users perceive high-frequency idioms as more transparent, seems less likely, because the adult frequencies in Study 1, which are retrieved from the 700-million-word Lassy Large corpus (Van Noord et al. 2013), do not show the same effect.

In our second study, we attempted to fill the gap with respect to idiom knowledge in young children. Our aim was to test whether children's familiarity increased with idiom exposure. We found that the familiarity ratings of the young children (age 7) were not influenced by idiom frequency. However, the familiarity ratings of older children (from age 9) increased significantly with increasing frequencies from the children's books corpus, while no effect of the adult frequencies was found. The results of the nine-year-old children signal that a reliable estimate of children's exposure is necessary for measuring an influence of the frequency of occurrence. The frequencies from the children's books corpus may not have been a good estimate of the idiom exposure of seven-year-old children, because less than 9% of the corpus contained books that were suited for children younger than seven years old (see Table 1). That is, similarly to the adult frequencies not being representative of nine-year-olds idiom exposure, idiom frequencies from the whole children's books corpus may not be representative of the youngest children's exposure. In addition, our idiom list, which was originally created for research in adults, may not have contained enough idioms that were familiar to these younger children, as discussed before.

Besides frequency, we investigated the influence of decomposability on idiom knowledge in children. We see that the familiarity ratings of the older children are influenced by decomposability, even though the underlying decomposability ratings from Sprenger et al. (2019) were provided by (young) adult participants in the age range 21–26 years, and so may not be representative for children's perception of the idioms' decomposability. The aforementioned study by Carrol (2023) reports that transparency does not change with age, and we seem to see at least some of that effect in our data as well.

The absence of an effect of decomposability on the familiarity ratings of the young children (age 7) in our study may be surprising in the light of previous studies that have found effects of decomposability in seven-old children's interpretation of idioms without context (e.g., Cain et al. 2009; Gibbs 1991). One of the reasons is that our study asked children to rate their familiarity with the idioms, rather than asking them to select the idioms interpretations. Because the list of idioms was not representative for young children, the amount of familiar idioms may not been sufficient to show an effect of decomposability. In addition, we investigated decomposability as a continuum instead of a categorical predictor (i.e., comparing highly decomposable idioms with non-decomposable idioms), which requires more observations for finding a significant trend. An alternative explanation is that the youngest age group experiences more difficulties with recognizing idioms without context, because their figurative competence skills are not sufficiently developed.

Our third study aimed to investigate the familiarity ratings of older children in more detail. We had expected to find stronger effects of frequency and decomposability with increasing age, but the limited number of observations per participant and item reduced the power of the effects, so that we did not find any differences between age groups. However, the overall results in Study 3 are quite similar to our findings for the older children in Study 2: the average familiarity rating for the Grade 6 (9 years old) children in Study 2 and Study 3 is highly similar (0.341 in Study 2 vs. 0.354 in Study 3; see Figure 3). Interestingly, we do see a numerical increase in average familiarity ratings with age: 0.354 for Grade 6, 0.428 for Grade 7, and 0.483 for Grade 8, but this trend is not significant. In addition, the effects of frequency and decomposability—while much weaker in Study 3 than in Study 2—go in the same direction as the results for the older children in Study 2. There are significant trends for decomposability and child frequency, but no effect of adult frequency,

and higher familiarity ratings for idioms that appeared in the children's book corpus. In other words, the ratings obtained in Study 3 support the idea that children are exposed to more transparent idioms than adults, and that the frequency with which these idioms occur predicts idiom knowledge in older (9+ years) children.

The consistent effect of decomposability on familiarity ratings in older children and adults is in line with the findings of Sprenger et al. (2019), who reported that young adults provide higher familiarity ratings for idioms with higher decomposability. Adults older than 40 in their study did not show this effect. Sprenger et al. (2019) observed that the decomposability did not affect familiarity ratings once an idiom had been acquired and was highly familiar. The reason for this effect of decomposability in the current study and in the earlier study may be that decomposable idioms are more easily recognized as being idiomatic—and their idiomatic meaning more easily analyzed—than opaque idioms, and hence, they are perceived as potentially familiar. Study 1 and the study of Carrol (2023) provide an additional hypothesis: transparent or decomposable idioms are more frequently used in edited texts (including children's books) than opaque idioms. Maybe this effect of decomposability is, therefore, an indirect effect of idiom exposure.

6. Conclusions

In the present paper, we present three different approaches to the question of how idiom frequency and decomposability jointly contribute to idiom knowledge in children. In Study 1, we show that idioms do occur in Dutch children's literature, but that the set of idioms is likely to be different from that in adult language. In line with this observation, adult frequencies are not predictive of the frequency with which an item occurs in the children's book corpus. Also, we observe a strong preference for idioms that score high on decomposability. With an increase in books' target age, however, we also find more of the idioms from our adult sample, suggesting that the type and range of idioms that children are exposed to changes with age. Study 2 confirms our findings on the difference between idioms in children's books and adult language, showing that older children's familiarity ratings are predicted by frequencies in the former, but not the latter. Finally, Study 3 confirmed our observation that children are generally exposed to more decomposable idioms than adults, and that the frequency with which these idioms occur predicts idiom knowledge in older children. Taken together, our findings are in line with theoretical approaches that attribute an important role to the development of figurative competence for the development of idiom knowledge, as they stress the importance of idiom decomposability for children of nine years and older. In addition, our findings expose a need for better corpora of idiom use in children's and child-directed language, as well as detailed studies on idiom knowledge and understanding in young and older children.

suggestions in putting together the children's books corpus. *Study 2:* The authors thank the children, parents, and teachers of the primary school for participating in the experiment in Study 2. *Study 3:* The authors thank the Dutch National Weekend of Science (het Weekend van de Wetenschap) for their collaboration and advice in the project *Maak dat de kat wijs*. In addition, they thank Remco Wouts for the technical support, the Scholierenacademie of the University of Groningen for their collaboration in creating the educational materials and advertising our questionnaires at primary schools, and all anonymous teachers and children who participated in the online study.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Study 1
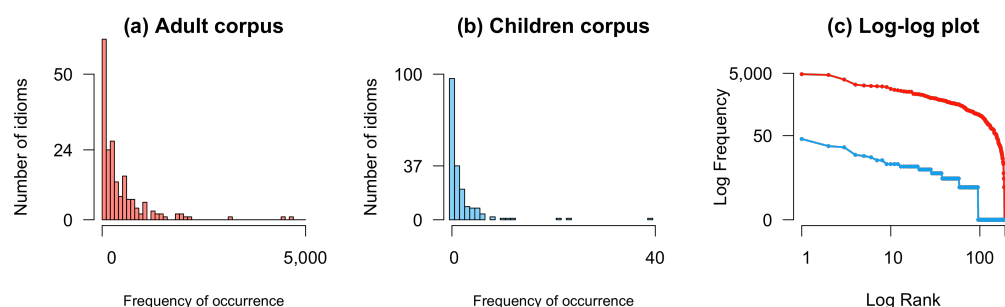
Appendix A.1. Figures with Extra Information

Figure A1. Frequency counts: (**a**) based on the Lassy Large corpus (Van Noord et al. 2013) and (**b**) based on the children's books corpus. The x-axis shows the (binned) frequency counts, and the y-axis shows the number of idioms with a frequency count in that bin. Panel (**c**) compares the two distributions by means of a log-log plot of the sorted frequency counts, with on the x-axis the log-transformed rank, and on the y-axis the log-transformed frequency count.
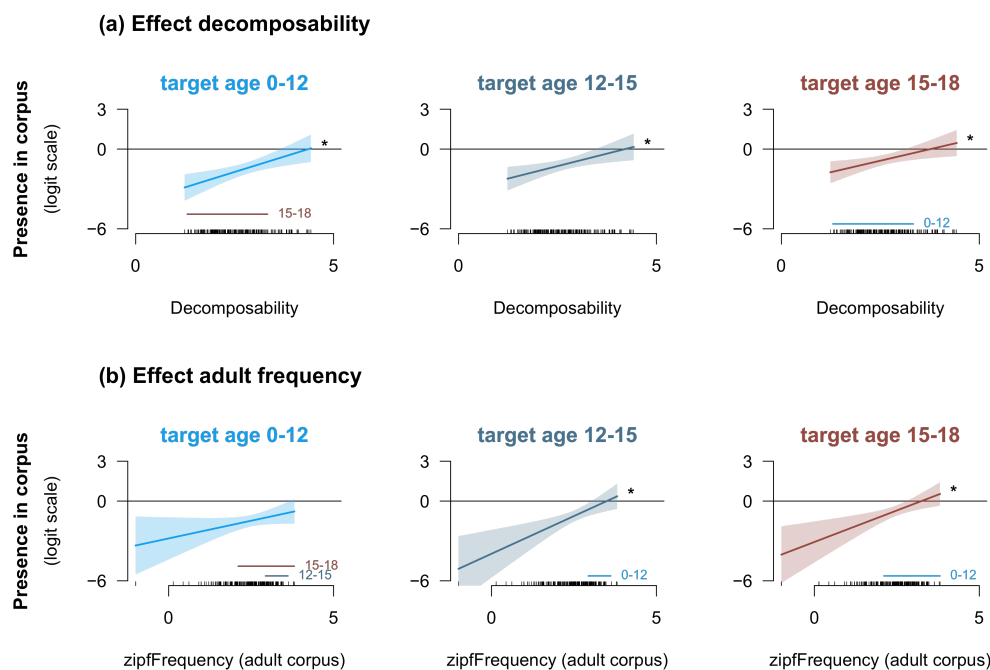

Figure A2. Comparison of the effects of decomposability (top row, panel (**a**)) and adult frequency (bottom row, panel (**b**)) on the three subcollections with target ages 0–12 (left), 12–15 (center), and 15–18 (right). *Panel (a):* the x-axis represents the decomposability scale, and the y-axis the estimated

log odds (logit scale) of whether or not the idioms appear in the children's corpus. *Panel (b):* the x-axis represents the log-transformed frequency counts based on the adult corpus, and the y-axis the estimated log odds (logit scale) of whether or not the idioms appear in the children's corpus. The symbol '*' indicates a significant trend for this specific age group. Horizontal lines indicate significant differences with other age groups.

Appendix A.2. List of Books in the Children's Books Corpus

Table A1. Books 0–6 years. Library codes: AB, AP, and AK (baby, toddlers, and preschoolers). In this category, we chose books with more text than average.

Nr	Author (Year)	Title
1.	Lizzie Finlay (2020): ISBN: 9789053417799	De Woeste zoete Wolf.
2.	Rachel Rooney (2020): ISBN: 9789053417492	Het probleem van Problemen.
3.	Tjibbe Veldkamp (2004): ISBN: 9789000035588	Tim op de Tegels.
4.	Ingrid en Dieter Schubert (1992): ISBN: 9789060698419	Woeste Willem.
5.	Martin Waddell en Barbara Firth (1988): ISBN: 9789047707646	Welterusten Kleine Beer.
6.	Marius van Dokkum (2007): ISBN: 9789072736543	Opa Jan wint een Olifant.

Table A2. Books 6–9 years. Library codes: A.

Nr	Author (Year)	Title
7.	Hans Hagen (2020): ISBN: 9789045125169	De mooiste Jubelientjes.
8.	Jochem Myer (2015): ISBN: 9789025867898	De Gorgels.
9.	Jochem Myer (2018): ISBN: 9789025875350	De Gorgels en het geheim van de Gletsjer.
10.	Annie M. G. Schmidt (1950s): ISBN: 9789045102252	Jip en Janneke.
11.	Guus Kuijer (1975): ISBN: 9789021432625	Met de Poppen Gooien.
12.	Mirjam Oldehave (2006): ISBN: 9789021680149	Mees Kees - een pittig klasje.
13.	Paul van Loon (2005): ISBN: 9789025864477	Boze Drieling.
14.	Janneke Schotveld (2019): ISBN: 9789000369263	Het Kattenmannetje en andere sprookjes.
15.	Dick Laan (1939): ISBN: 9789047509721	De Avonturen van Pinkeltje.
16.	Hanna Kraan (1990): ISBN: 9789060697924	Verhalen van de Boze Heks.
17.	Manon Sikkel (2016): ISBN: 9789024574865	Geheim agent oma.

Table A3. Books 9–12 years. Library codes: B.

Nr	Author (Year)	Title
18.	Paul van Loon (1991): ISBN: 9789025871406	De Griezelbus.
19.	Piet Prins (1954): ISBN: 9789060154861	Snuf de Hond.
20.	Hotze de Roos (1949): ISBN: 9789020667011	De schippers van de Kameleon.
21.	Annet Schaap (2017): ISBN: 9789045120379	Lampje.
22.	Jacques Vriens (1999): ISBN: 902699227	Achtste-Groepers huilen niet.
23.	Elisabetta Dami (2017): ISBN: 9789085924302	Stilton Ridder voor een dag.
24.	Tonke Dragt (1962): ISBN: 9789025868444	De brief voor de koning.
25.	Tonke Dragt (1966): ISBN: 9025833985	De Zevensprong.
26.	Francine Oomen (1998): ISBN: 9789026995590	Hoe overleef ik mijn vakantie.
27.	J. K. Rowling (1998): ISBN: 9076174083	Harry Potter en de Steen der Wijzen.
28.	John Flanagan (2004): ISBN: 9789025742843	De Ruïnes van Gorlan.

Table A4. Books 12–15 years. Library codes: C.

Nr	Author (Year)	Title
29.	Thea Beckman (1983): ISBN: 9789060695401	Hasse Simonsdochter.
30.	Carry Slee (1996): ISBN: 9789048849178	Spijt.
31.	Jan Terlouw (1971): ISBN: 9060690885	De Koning van Katoren.
32.	Jan Terlouw (1972): ISBN: 9789060691182	Oorlogswinter.
33.	Thea Beckman (1973): ISBN: 9789060691670	Kruistocht in Spijkerbroek.
34.	Anne Frank (1947): ISBN: 9789035133068	Het Achterhuis.
35.	Roald Dahl (1978): ISBN: 9789026120763	Hendrik Meier.
36.	Anna van Praag (2021): ISBN: 9789047712534	Noorderlicht.
37.	Dolf Verroen (2019): ISBN: 9789025878238	Niemand die het ziet.
38.	J.K. Rowling (2000): ISBN: 9789076174198	Harry Potter en de Vuurbeker.
39.	Johan Fabricius (1924): ISBN: 9789025834609	De Scheepsjongens van Bontekoe.

Table A5. Books 15–18 years. Library codes: D Young adults.

Nr	Author (Year)	Title
40.	J. K. Rowling (2005): ISBN: 9061697662	Harry Potter en de Halfbloedprins.
41.	Beau Charlotte (2020): ISBN: 9789044839159	Als ik er niet meer ben.
42.	Rindert Kromhout (2013): ISBN: 9789025864071	April is de wreedste maand.
43.	John Green (2012): ISBN: 9789047706618	Een weeffout in onze sterren.
44.	Thomas Olde Heuvelt (2013): ISBN: 9789024573349	Hex.
45.	Aidan Chambers (1985): ISBN: 9789045125770	Je moet dansen op mijn graf.
46.	John Boyne (2006): ISBN: 9789022568705	De jongen in de gestreepte pyjama.
47.	Wiliam Golding (1954, 1960, 1962): ISBN: 9789025317522	Heer der Vliegen.
48.	Herman Koch (2009): ISBN: 9789041413680	Het diner.
49.	Tommy Wieringa (2005): ISBN: 9789023455493	Joe Speedboot.
50.	Robert Vuijsje (2008): ISBN: 9789038890616	Alleen maar nette mensen.

Appendix A.3. Most Frequent Idioms in Children's Books Corpus

Table A6. Most frequent idioms in children's books corpus.

| Idiom	Meaning	Child Frequency	Adult Frequency
'Toen leerde hij het uit het hoofd.' *Then learned he it out the head.*	to learn something by heart	39	688
'Toen maakte hij zich uit de voeten.' *Then made he himself out the feet.*	to flee	23	1020
'Toen liep hij in de val.' *Then walked he in the trap.*	to walk into a trap	21	3153
'Toen vatte hij hem in de kraag.' *Then caught he him in the collar.*	to arrest someone	12	557
'Toen stond hij met zijn rug tegen de muur.' *Then stood he with his back against the wall.*	to have no way out	11	670
'Toen kreeg hij het onder de knie.' *Then got he it under the knee.*	to get the hang of something	10	1091
'Toen hield hij een oogje in het zeil.' *Then he held an eye in the sail.*	to keep an eye on things	8	1281
'Toen deed hij het in zijn broek.' *Then did he it in his pants.*	to be afraid	8	662
'Toen viel hij met de deur in huis.' *Then fell he with the door in house.*	to come straight to the point	6	1424
'Toen kwam hij uit de kast.' *Then came he out the closet.*	to openly admit one's homosexual nature for the first time	6	597
'Toen streek hij met de eer.' *Then brushed he with the honour.*	to take the credit you do not deserve	6	167
'Toen klopte zijn hart in zijn keel.' *Then beat their heart in their throat.*	to be afraid	6	88

Appendix B. Study 2
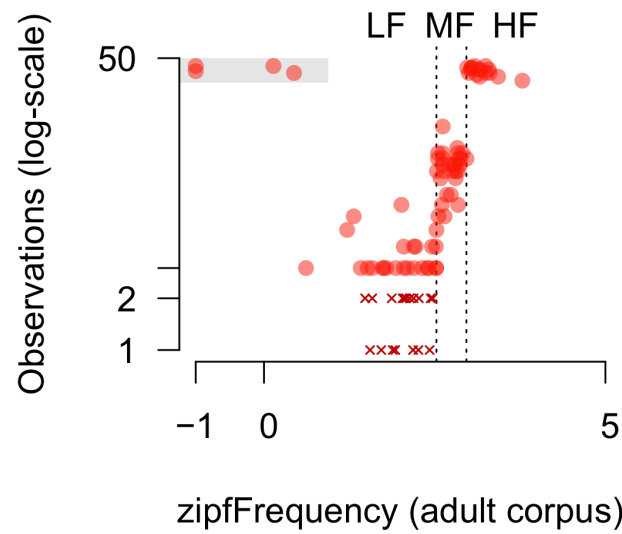
Figure A3. Number of observations by frequency per idiom. Idioms indicated with 'x' have been excluded from analysis (one or two observations). Idioms in gray box are control idioms.

Appendix B.1. Most Familiar Idioms

List of idioms that were rated as being familiar by more than three children and the overall familiarity rating was more than 0.5.

Table A7. Grade 4.

Idiom	Meaning	Familiar	Unfamiliar
'Toen hield hij een oogje in het zeil.' *Then he held an eye in the sail.*	To keep an eye on things	8	5

Table A8. Grade 6.

Idiom	Meaning	Familiar	Unfamiliar
'Toen hield hij een oogje in het zeil.' *Then he held an eye in the sail.*	to keep an eye on things	15	0
'Toen sprong hij een gat in de lucht.' *Then jumped he a whole in the air.*	to jump for joy	13	1
'Toen viel hij met de deur in huis.' *Then fell he with the door in house.*	to come straight to the point	10	4
'Toen bleef hij met beide benen op de grond.' *Then stayed he with both legs on the ground.*	to have one's feet firmly on the ground	5	0
'Toen had hij een appeltje voor de dorst.' *Then had he an appel for the thirst.*	to have a buffer	4	0
'Toen sloeg hij de spijker op de kop.' *Then hit he the nail on the head.*	to hit the nail on the head	4	0

Notes

1    $zipf = \log_{10}(\frac{f_i * 1000000}{N}) + \log_{10}(1000)$.

2   The Dutch title of the project is a well-known Dutch idiom with the literal translation *Teach that to the cat*, meaning 'I do not believe one bit of it!'.

3   Note that participants entered their age in years, therefore we do not have the precision to provide the ages in months.

References

Arnon, Inbal, and Morten H. Christiansen. 2017. The role of multiword building blocks in explaining L1–L2 differences. *Topics in Cognitive Science* 9: 621–36. [CrossRef] [PubMed]

Arnon, Inbal, and Neal Snider. 2010. More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language* 62: 67–82. [CrossRef]

Arnon, Inbal, and Uriel Cohen Priva. 2013. More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech* 56: 349–71. [CrossRef] [PubMed]

Ayto, John. 2010. *Oxford Dictionary of English Idioms*. Oxford: Oxford University Press. [CrossRef]

Bannard, Colin, and Danielle Matthews. 2008. Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science* 19: 241–48. [CrossRef] [PubMed]

Barlow, Michael. 2000. Usage, blends and grammar. In *Usage-Based Models of Language*. Edited by M. Barlow and S. Kemmer. Stanford: CSLI Publications, pp. 315–45.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. 1999. *Longman Grammar of Spoken and Written English.* Harlow: Longman.

Bonin, Patrick, Alain Méot, and Aurelia Bugaiska. 2013. Norms and comprehension times for 305 French idiomatic expressions. *Behavior Research Methods* 45: 1259–71. [CrossRef]

Brysbaert, Marc, Michaël Stevens, Paweł Mandera, and Emmanuel Keuleers. 2016. How many words do we know? practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology* 7: 1116. [CrossRef]

Cacciari, Cristina, and Patrizia Tabossi. 1988. The comprehension of idioms. *Journal of Memory and Language* 27: 668–83. [CrossRef]

Cain, Kate, Andrea S. Towse, and Rachael S. Knight. 2009. The development of idiom comprehension: An investigation of semantic and contextual processing skills. *Journal of Experimental Child Psychology* 102: 280–98. [CrossRef]

Carrol, Gareth. 2023. Old Dogs and New tricks: Assessing Idiom Knowledge Amongst Native Speakers of Different Ages. *Journal of Psycholinguistic Research* 52: 2287–302. [CrossRef]

Carrol, Gareth, Jeannette Littlemore, and Margaret Gillon Dowens. 2018. Of false friends and familiar foes: Comparing native and non-native understanding of figurative phrases. *Lingua* 204: 21–44. [CrossRef]

Cieślicka, Anna. 2006. Literal salience in on-line processing of idiomatic expressions by second language learners. *Second Language Research* 22: 115–44. [CrossRef]

Cieślicka, Anna B. 2015. Idiom acquisition and processing by second/foreign language learners. In *Bilingual Figurative Language Processing*. Edited by R. M. Heredia and A. B. Cieślicka. Cambridge: Cambridge University Press, pp. 208–44. [CrossRef]

Conklin, Kathy, and Norbert Schmitt. 2008. Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics* 29: 72–89. [CrossRef]

Conklin, Kathy, and Norbert Schmitt. 2012. The processing of formulaic language. *Annual Review of Applied Linguistics* 32: 45–61. [CrossRef]

Contreras Kallens, Pablo, and Morten H. Christiansen. 2022. Models of language and multiword expressions. *Frontiers in Artificial Intelligence* 5: 781962. [CrossRef]

Davies, Mark. 2016. Corpus of News on the Web (NOW). Available online: https://www.english-corpora.org/now/ (accessed on 1 October 2024).

de Groot, Hans. 1999. *Van Dale Idioomwoordenboek: Verklaring en herkomst van uitdrukkingen en gezegden*. Utrecht and Antwerp: Van Dale Lexicografie.

Erman, Britt, and Beatrice Warren. 2000. The idiom principle and the open choice principle. *Text & Talk* 20: 29–62. [CrossRef]

Gibbs, Raymond W. 1987. Linguistic factors in children's understanding of idioms. *Journal of Child Language* 14: 569–86. [CrossRef]

Gibbs, Raymond W. 1991. Semantic analyzability in children's understanding of idioms. *Journal of Speech, Language, and Hearing Research* 34: 613–20. [CrossRef]

Hasties, Trevor J., and Robert J. Tibshirani. 1990. *Generalized Additive Models*. London: Chapman and Hall.

Hubers, Ferdy, Catia Cucchiarini, Helmer Strik, and Ton Dijkstra. 2019. Normative data of dutch idiomatic expressions: Subjective judgments you can bank on. *Frontiers in Psychology* 10: 1075. [CrossRef]

Keysar, Boaz, and Bridget Bly. 1995. Intuitions of the transparency of idioms: Can one keep a secret by spilling the beans? *Journal of Memory and Language* 34: 89–109. [CrossRef]

Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý, and Vít Suchomel. 2014. The Sketch Engine: Ten years on. *Lexicography* 1: 7–36. [CrossRef]

Kuiper, Koenraad, Georgie Columbus, and Norbert Schmitt. 2009. The acquisition of phrasal vocabulary. In *Language Acquisition*. Edited by S. Foster-Cohen. London: Palgrave Macmillan, pp. 216–40. [CrossRef]

Levorato, Maria Chiara, and Cristina Cacciari. 1992. Children's comprehension and production of idioms: The role of context and familiarity. *Journal of Child Language* 19: 415–33. [CrossRef] [PubMed]

Levorato, M. Chiara, and Cristina Cacciari. 1995. The effects of different tasks on the comprehension and production of idioms in children. *Journal of Experimental Child Psychology* 60: 261–83. [CrossRef] [PubMed]

Lin, Xihong, and Daowen Zhang. 1999. Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 61: 381–400. [CrossRef]

Martinez, Ron, and Norbert Schmitt. 2012. A phrasal expressions list. *Applied Linguistics* 33: 299–320. [CrossRef]

Mathôt, Sebastiaan, Daniel Schreij, and Jan Theeuwes. 2012. OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods* 44: 314–24. [CrossRef]

Moon, Rosamund. 1998. *Fixed Expressions and Idioms in English: A Corpus-Based Approach*. Oxford: Oxford University Press. [CrossRef]

Nicoladis, Elena. 2019. 'I have three years old': Cross-linguistic Influence of Fixed Expressions in a Bilingual Child. *Journal of Monolingual and Bilingual Speech* 1: 80–93. [CrossRef]

Nippold, Marilyn A., and Catherine L. Taylor. 1995. Idiom understanding in youth: Further examination of familiarity and transparency. *Journal of Speech, Language, and Hearing Research* 38: 426–33. [CrossRef]

Nippold, Marilyn A., and Mishelle Rudzinski. 1993. Familiarity and transparency in idiom explanation: A developmental study of children and adolescents. *Journal of Speech, Language, and Hearing Research* 36: 728–37. [CrossRef]

Nippold, Marilyn A., and Stephanie Tarrant Martin. 1989. Idiom interpretation in isolation versus context: A developmental study with adolescents. *Journal of Speech, Language, and Hearing Research* 32: 59–66. [CrossRef]

Nordmann, Emily, Alexandra A. Cleland, and Rebecca Bull. 2014. Familiarity breeds dissent: Reliability analyses for British-English idioms on measures of familiarity, meaning, literality, and decomposability. *Acta Psychologica* 149: 87–95. [CrossRef]

Nunberg, Geoffrey, Ivan A. Sag, and Thomas Wasow. 1994. Idioms. *Language* 70: 491–538. [CrossRef]

Pawley, Andrew, and Frances Hodgetts Syder. 1983. Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In *Language and Communication*. Edited by J. C. Richards and R. W. Schmidt. London: Routledge, pp. 191–226.

Pinker, Steven. 1991. Rules of language. *Science* 253: 530–35. [CrossRef] [PubMed]

Piquer-Píriz, Ana M. 2020. Figurative language and young L2 learners. In *Metaphor in Foreign Language Instruction*. Edited by A. M. Piquer-Píriz and R. Alejo-González. Berlin and Boston: De Gruyter Mouton, pp. 57–78. [CrossRef]

Reuterskiöld, Christina, and Diana Van Lancker Sidtis. 2013. Retention of idioms following one-time exposure. *Child Language Teaching and Therapy* 29: 219–31. [CrossRef]

Sprenger, Simone, Amélie la Roi, and Jacolien van Rij. 2019. The development of idiom knowledge across the lifespan. *Frontiers in Communication* 4: 29. [CrossRef]

Sprenger, Simone, and Jacolien van Rij. 2019. *Maak dat de kat wijs*. Citizen Science Project Organized with the National Dutch Weekend of Science (Weekend van de Wetenschap) in 2019. Available online: https://maakdatdekatwijs.nl (accessed on 1 October 2024).

Swinney, David A., and Anne Cutler. 1979. The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior* 18: 523–34. [CrossRef]

Tabossi, Patrizia, Lisa Arduino, and Rachele Fanari. 2011. Descriptive norms for 245 Italian idiomatic expressions. *Behavior Research Methods* 43: 110–23. [CrossRef]

Titone, Debra A., and Cynthia M. Connine. 1994. Descriptive norms for 171 idiomatic expressions: Familiarity, compositionality, predictability, and literality. *Metaphor and Symbol* 9: 247–70. [CrossRef]

Titone, Debra A., and Cynthia M. Connine. 1999. On the compositional and noncompositional nature of idiomatic expressions. *Journal of Pragmatics* 31: 1655–74. [CrossRef]

Tremblay, Antoine, and Benjamin V. Tucker. 2011. The effects of N-gram probabilistic measures on the recognition and production of four-word sequences. *The Mental Lexicon* 6: 302–24. [CrossRef]

Tremblay, Antoine, Bruce Derwing, Gary Libben, and Chris Westbury. 2011. Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning* 61: 569–613. [CrossRef]

Van Heuven, Walter J.B., Pawel Mandera, Emmanuel Keuleers, and Marc Brysbaert. 2014. Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology* 67: 1176–90. [CrossRef] [PubMed]

Van Noord, Gertjan, Gosse Bouma, Frank van Eynde, Daniel de Kok, Jelmer van der Linde, Ineke Schuurman, Erik Tjong Kim Sang, and Vincent Vandeghinste. 2013. Large scale syntactic annotation of written dutch: Lassy. In *Essential Speech and Language Technology for Dutch: Results by the STEVIN Programme*. Edited by P. Spyns and J. Odijk. Berlin: Springer, pp. 147–64. [CrossRef]

van Rij, Jacolien, Martijn Wieling, R. Harald Baayen, and Hedderik van Rijn. 2022. itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs. R package version 2.4.1. Available online: https://cran.r-project.org/package=itsadug (accessed on 1 October 2024).

van Rij, Jacolien, Petra Hendriks, Hedderik van Rijn, R. Harald Baayen, and Simon N. Wood. 2019. Analyzing the time course of pupillometric data. *Trends in Hearing* 23: 1–22. [CrossRef]

Wieling, Martijn. 2018. Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics* 70: 86–116. [CrossRef]

Winner, Ellen. 1988. *The Point of Words: Children's Understanding of Metaphor and Irony*. Harvard: Harvard University Press.

Wood, Simon N. 2017. *Generalized Additive Models: An Introduction with R.* London: Chapman and Hall/CRC. [CrossRef]

Wray, Alison. 2002. *Formulaic Language and the Lexicon.* Cambridge: Cambridge University Press. [CrossRef]

Zurer-Pearson, Barbara. 1990. The comprehension of metaphor by preschool children. *Journal of Child Language* 17: 185–203. 10.1017/s0305000900013179. [CrossRef] [PubMed]