




ORIGINAL ARTICLE

# French-speaking teenagers' mastery of connectives: the role of vocabulary size and exposure to print

Ekaterina Tskhovrebova<sup>1,\*</sup>, Sandrine Zufferey<sup>1</sup> and Elena Tribushinina<sup>2</sup>

<sup>1</sup>Department of French Language and Literature, University of Bern, Bern, Switzerland and <sup>2</sup>Department of Languages, Literature and Communication, Utrecht University, Utrecht, Netherlands

\*Corresponding author. Email: [ekaterina.tskhovrebova@unibe.ch](mailto:ekaterina.tskhovrebova@unibe.ch)

(Received 25 February 2022; revised 12 July 2022; accepted 19 September 2022; first published online 28 October 2022)

## Abstract

Connectives such as *however* and *since* play an important role for marking coherence relations in discourse and therefore are crucial for reading comprehension, which in turn is a strong predictor of academic success. Most research on the acquisition of connectives targeted younger children. Yet there is evidence that connective development extends well into adolescence and even adult speakers have difficulties with some coherence relations when they are conveyed by infrequent connectives bound to the written mode. In this paper, we tested the use of connectives encoding different coherence relations and bound to either the oral or the written modes. We studied the performance of native French-speaking teenagers ( $N = 154$ ,  $M_{\text{age}} = 14.43$ , range: 12–19) in a cloze task and also assessed whether teenagers' vocabulary level and degree of exposure to print predicted the accuracy of connective use. Our findings show that the ability to use connectives appropriately increases with age. However, age played a lesser role compared to vocabulary knowledge and degree of exposure to print, thus indicating that lexicon size and reading habits are important factors explaining individual differences in the acquisition of connectives.

**Keywords:** discourse connectives; coherence relations; teenagers; exposure to print; vocabulary size

Connectives are linguistic devices signaling coherence relations such as causality and contrast between discourse segments (e.g., Halliday & Hasan, 1976). In previous research, connectives were found to play a crucial role for the comprehension of coherence relations and, more generally, reading comprehension (e.g., Degand & Sanders, 2002; Millis et al., 1993). More cohesive texts are better understood (van Silfhout et al., 2015), especially by less proficient readers (Linderholm et al., 2000; van Silfhout, 2014) and by speakers with a limited knowledge of a subject (Ozuru et al., 2009) or individuals with a language impairment (e.g., Corkett et al., 2006). Reading comprehension, in turn, is an important predictor of academic

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial licence (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use.

success in various subject areas such as history (e.g., Beek, 2020), mathematics (e.g., Fuentes, 1998; Jordan *et al.*, 2006; Salihu *et al.*, 2018), and sciences in general (e.g., Akbaşlı *et al.*, 2016; Imam *et al.*, 2014; Korpershoek *et al.*, 2015; O'Reilly & McNamara, 2007). However, more than ten million 15-year-old students from all over the world, who participated in the Programme for International Student Assessment in 2018, had difficulties with reading, as they were unable to complete even the most basic reading tasks (Schleicher, 2019). Considering that connectives importantly contribute to the comprehension of texts (see, e.g., Degand & Sanders, 2002; Millis & Just, 1994) and that they are an essential part of basic academic language skills (Barr *et al.*, 2019; RAND Reading Study Group & Snow, 2002), there is an urge to unravel factors explaining individual differences in connective knowledge. The majority of studies examining use and comprehension of connectives either provided evidence for the mastery of connectives by adults (e.g., Canestrelli *et al.*, 2013; Zufferey & Gygax, 2020a) or examined the mechanisms of their acquisition in young children (e.g., Cain & Nash, 2011; Peterson, 1986). They showed, for instance, that the most important predictors of appropriate connective use for children were age as well as the degree of complexity of the coherence relations (e.g., Cain & Nash, 2011; Evers-Vermeul & Sanders, 2009). In contrast, relatively little is known about connective acquisition by older children and teenagers, which is surprising because speakers of this age are also exposed to connectives on a regular basis, not only in texts related to their language classes but also in texts used for other school subjects. For instance, passages (1)–(5) illustrate several occurrences of connectives in maths and history textbooks, used by the 10<sup>th</sup> grade students (13–14 years) in the French-speaking part of Switzerland. In examples (1)–(3), connectives highlight causal and consequence relations between sentences, drawing readers' attention to how these relations should be interpreted.

- (1) Le prix à payer et la quantité d'essence sont proportionnels. *En effet*, pour obtenir le prix à payer on multiplie la quantité d'essence par le prix d'un litre. "The price to be paid and the quantity of petrol are proportional. For this reason, to obtain the price to be paid, the quantity of petrol is multiplied by the price of a litre."
- (2) Le facteur commun à chaque monôme est  $5 \times$ . On peut *donc* le mettre en évidence. "The common factor for each monomial is  $5 \times$ . We can therefore highlight it."
- (3) Durant cette période de troubles, un général, Napoléon Bonaparte, stabilise la situation. Il passe *ainsi* pour l'homme providentiel et en profite pour prendre de plus en plus de pouvoir. "During this period of unrest, a general, Napoleon Bonaparte, stabilises the situation. He is therefore seen as the providential man and takes advantage of this to gain more and more power."

In comparison, examples (4) and (5) include connectives that not only highlight contrastive and concessive relations between sentences but are also crucial for appropriate understanding of these coherence relations. In other words, not knowing these connectives would completely impede the understanding of the meaning

of the whole passage, as these relations are difficult to infer when they are not signaled explicitly by a connective.

(4) De nombreux États, par exemple, prennent en charge l'assistance aux pauvres, tâche accomplie jusque-là uniquement par l'Église. *Toutefois*, cette aide peut s'accompagner de l'enfermement ou de la mise au travail forcé.

"Many states, for example, are taking over assistance to the poor, a task previously performed only by the church. However, this assistance may be accompanied by confinement or forced labour."

(5) *Bien que* les cantons suisses ne combattent pas directement dans la guerre de Trente Ans, ils sont mentionnés dans les Traités de Westphalie.

"Although the Swiss cantons did not fight directly in the Thirty Years' War, they are mentioned in the Treaties of Westphalia."

Although several studies have included teenagers in their experiments (Kleijn et al., 2019; McClure & Geva, 1983; Nippold et al., 1992; Van Silfhout et al., 2015; Zufferey & Gygax, 2020b), a number of open research questions remain about the way connectives keep on developing during this period, and especially why some connectives seem to be more challenging than others in this age group. We will tackle this question in this paper because mastery of connectives is important for reading comprehension (e.g., Traxler et al., 1997) and is an integral part of core academic language skills (Barr et al., 2019; Snow & Uccelli, 2009). Furthermore, a full-fledged acquisition of connectives is crucial for the development of adult-level mastery of language (Berman, 2004). To investigate whether the competence with connectives is modulated by teenager individual characteristics, our study also included two background measures of individual difference, namely vocabulary knowledge and exposure to print.

### Mastery of connectives during teenage years

In early teenage years (around the age of 12), pupils are able to understand the main types of coherence relations such as causality, contrast, concession, and addition (e.g., Crosson & Lesaux, 2013; McClure & Geva, 1983; Nippold et al., 1992; Zufferey & Gygax, 2020b). For example, McClure and Geva (1983), who studied the use of adversative connectives *but* and *although* in a cloze task, concluded that teenagers master both connectives well by age 9 (fourth grade). Nippold et al. (1992) also found that, by age 12, teenagers mastered equally well connectives encoding relations of addition (e.g., *moreover*), consequence (e.g., *consequently*), concession (e.g., *however*), and contrast (e.g., *contrastively*), as measured by both sentence continuation and cloze tasks.

Not all the connectives expressing a particular coherence relation are always used correctly by teenagers in this age group. There is evidence that the difficulty of certain connectives may stem not only from the complexity of the coherence relation they encode (Sanders et al., 1992) but from other factors as well. Crosson and Lesaux (2013) focussed on four coherence relations (additive, temporal, adversative, and causal), represented by connectives with different degrees of familiarity. Degrees

of familiarity were attributed to connectives depending on the proportion of children who knew them in a given age group. For instance, for causal relations, they tested the connectives *because*, *therefore*, *consequently*, and *hence*. In this sample, *because* had the highest degree of familiarity and *hence* the lowest. Using a cloze sentence task, the authors found that young teenagers performed better with connectives that had a higher degree of familiarity, irrespective of the type of coherence relation that these connectives encoded.

In addition, the effect of familiarity may be intertwined with that of mode, that is, whether connectives are typically used in spoken or written language. Indeed, children are exposed to oral speech starting from birth, while extensive exposure to written language comes much later. Children start to be exposed to writing mostly through schooling, and this exposure becomes extensive only in secondary school when teenagers become truly autonomous readers of various text genres (Nippold, 2004, 2008). Consequently, it is plausible to assume that connectives that are mostly used in writing are mastered less well than those bound to oral speech. There are studies that operationalized degree of connective familiarity through their frequency in corpora (Nippold *et al.*, 1992; Tskhovrebova *et al.*, 2022; Zufferey & Gyga, 2020b). However, even in these studies, which disentangled the effects of mode and frequency by studying only the connectives common for the written mode, frequency was still found to be an important predictor of connective use (Nippold *et al.*, 1992; Tskhovrebova *et al.*, 2022; Zufferey & Gyga, 2020b).

These studies had several limitations, which may have influenced their outcomes. For example, Nippold and colleagues (1992) did not explicitly test the effect of frequency on teenagers' competence with connectives. The authors suggested that frequency was likely to be an important predictor only in their post hoc explanation of the results. This idea was later confirmed by other researchers who designed their experiments taking into account the effects of frequency (Tskhovrebova *et al.*, 2022; Zufferey & Gyga, 2020b). In both studies, the authors examined the usage of four French connectives that are mostly used in writing, encode different coherence relations, and have different frequencies in written corpora. The main difference between the two papers was that Zufferey and Gyga (2020b) studied the usage of connectives only in a sentence cloze test and focused on high-school students. In contrast, Tskhovrebova and colleagues (2022) explored a wider age range, including secondary-school students, and compared performance in a sentence cloze task with a more ecological text cloze task. Despite the methodological differences, the two studies yielded converging results: They found that teenagers perform worse with the less frequent written connectives *en outre* "in addition" and *aussi* "therefore" than with the more frequent connectives *en effet* "for" and *toutefois* "however."

In these two studies, however, the effect of frequency may have been intertwined with the effect of polyfunctionality, when a connective can signal different coherence relations depending on context, as two of the four connectives tested were polyfunctional. The importance of connective polyfunctionality as a factor lowering teenagers' ability to handle connectives was hinted at in the error analysis conducted by the authors. Indeed, it was observed that teenagers often erroneously used the connective *aussi* instead of connective *en outre*. While *en outre* is monofunctional and can only be used to signal additive relations (i.e., when new information is added to previous segment of text), *aussi* can be used as an additive connective along

with its consequential meaning and is therefore polyfunctional. The additive meaning of *aussi* is also by far its most frequent meaning in language use. It is therefore possible that teenagers may have followed the probabilistic approach to connective interpretation (Asr & Demberg, 2020) and inferred the more frequent additive function of *aussi*. However, this answer was erroneous, as *aussi* can be used in its additive function only in sentence-medial or final position (Roze et al., 2012), and in these studies connectives were missing only in sentence-initial position. It may therefore be important to rule out the effect of polyfunctionality in future studies with teenagers, since previous research shows that even for adults it may be challenging to judge the appropriate uses of polyfunctional connectives, especially when they are used in infrequent functions (e.g., Zufferey et al., 2015).

In addition, another limitation of these studies is that they examined a restricted number of connectives. In the present paper, we aim to fill in these gaps by analyzing a larger variety of connectives, encoding a greater number of coherence relations, from the oral and the written modes. To rule out the effect of polyfunctionality, we will examine only monofunctional connectives.

### Individual variation in the mastery of connectives by teenagers

The mastery of connectives not only depends on the factors related to properties of connectives but also varies according to individual characteristics of teenagers. Previous studies, for instance, found variation related to teenagers' age (e.g., Nippold et al., 1992), reading proficiency (Van Silfhout et al., 2015), and academic background operationalized as different educational tracks followed by the pupils (Tskhovrebova et al., 2022; Van Silfhout et al., 2015; Zufferey & Gygax, 2020b). Our study continues this line of research and aims to gain a better understanding of individual variation and to reveal other potential sources of individual differences. We know, for example, that vocabulary knowledge continues to develop during teenage years and varies between speakers with different social, personal, and life-experience backgrounds (Nation & Coxhead, 2021). It is plausible that speakers with a richer general vocabulary level also know more connectives since connectives constitute a specific domain of lexical knowledge (Crosson & Lesaux, 2013). Indeed, Wetzel and colleagues (2020) found that lexicon size was a strong predictor of the performance with connectives in a sentence cloze task by adult native and non-native speakers of French. However, it is not clear whether this effect also holds for teenagers, especially considering the particular status of connectives in the lexicon. According to the declarative-procedural (DP) model of language (Ullman, 2001), language acquisition and use are supported by two brain memory systems. Declarative memory underlies the acquisition and use of idiosyncratic elements, such as words and irregular morphology. Procedural memory supports the acquisition and use of cognitive routines, regular morphology, and (partly) phonology. We would like to propose that connectives are not an ordinary part of lexicon, as they express procedural rather than declarative meaning, in contrast to the majority of other lexical items (Blakemore, 2002; Wilson, 2011; Wilson & Sperber, 1993). There is psycholinguistic evidence that connectives give speakers processing instructions (Britton, 1994; Gernsbacher, 1997; Sanders & Spooren, 2007) and guide

them in the way they should relate parts of text. For example, the connective *therefore* in the sentence (6), rather than expressing a conceptual meaning, indicates how the relation between the two clauses should be interpreted. It signals that the first clause should be analyzed as a cause and the second one as a consequence of the described event.

(6) Mindy had gotten asthma; therefore, she could not give her lecture.

The claim that connectives function as processing instructions has been supported by experimental data. In a visual-world experiment, Koehne and Demberg (2013) revealed not only that concessive and causal connectives in German evoke different expectations (i.e., give different processing instructions) for the upcoming coherence relation but also that concessive connectives elicit slower predictions for the forthcoming content than causal ones. To put it differently, this experiment shows that connectives of different complexity require different processing times when guiding readers in creating a coherent continuation of a sentence.

Canestrelli *et al.* (2013) also showed in a series of eye-tracking experiments in Dutch that there was a delay in the processing of the subjective causal connective *want* “because” compared to the objective causal connective *omdat* “because.” Moreover, this effect was observed immediately after these causal connectives, meaning that they instantly trigger a representation of a causal relation, way before the end of the second clause. In other words, it was shown that the instruction about the causal relation, which should be expected in the next clause, appears as soon as a reader sees a causal connective. Repeated exposure to connectives in texts leads to entrenchment of a cognitive routine to relate discourse segments based on the processing instruction encoded in the specific connective. Utilizing connectives for discourse processing can thus be seen as part of proceduralized knowledge, which is largely automatized in experienced readers, just like riding a bike is automatized in experienced cyclists. Hence, connectives appear to occupy an intermediate position between declarative memory (being parts of the lexicon) and procedural memory (being processing instructions).

Finally, there is evidence that knowledge of connectives and general vocabulary knowledge do not exactly overlap. It was found that knowledge of connectives significantly contributes to the improvement of reading comprehension (in a second language) when other factors such as vocabulary knowledge, reading fluency, and metacognitive knowledge are controlled for (Crosson & Lesaux, 2013; Welie *et al.*, 2017). In consequence, assessing whether the width of general vocabulary in young speakers predicts the appropriate usage of connectives would shed light on their nature and contribute to psycholinguistic theories at a more general level. If connectives are indeed processing instructions, as suggested by theoretical and experimental research, vocabulary size should not be the only predictor of their appropriate use.

Another important predictor of the competence with connectives in teenagers may be amount of exposure to print, as it is through written texts that speakers are potentially exposed to a greater number of connectives, used with more precise functions (see, e.g., Crible & Cuenca, 2017). Indeed, exposure to print, as

measured by the author recognition test (ART) (Stanovich & West, 1989), has been shown to predict mastery of connectives in adults (Zufferey & Gygax, 2020a). It was also found in previous studies that the ART test predicts various other linguistic skills, such as vocabulary knowledge, word recognition, and general reading ability in both children and adults (e.g., Spear-Swerling et al., 2010; West et al., 1993). But to the best of our knowledge, this phenomenon was not studied with teenagers. It is especially important to study the relation between exposure to print and the competence to use connectives at this age since the novel insights may lead to solutions on how this competence can be improved during middle- and high-school years, for instance, through the increase in the amount of in- or out-of-class reading. Thus, testing exposure to print as predictor for the usage of connectives in the teenage population will be an important contribution to the research in this domain.

Lastly, we aim to determine how important the age factor is, in comparison to vocabulary level and exposure to print. Previous research on the acquisition of connectives by older children indicates that this predictor is less strong than teenagers' academic background (Tskhovrebova et al., 2022). In striking contrast, in studies involving primary school children, age was always found to be one of the strongest predictors of connective usage and comprehension (e.g., Blything et al., 2015; Cain & Nash, 2011; Pyykkönen & Järvikivi, 2012). The DP model of language (Ullman, 2001) could potentially explain why age may be a less strong predictor in teenage years. Since declarative memory improves with age, older children more easily learn new words. In contrast, procedural learning ability abates with age. As connectives may also be part of procedural knowledge, age may play a less important role in teenage years when procedural learning becomes more demanding.

### Research questions of the present study

In this paper, we aim to examine the factors that predict the correct usage of connectives by teenagers. To cover the gaps from previous research, we will study six types of coherence relations, conveyed by monofunctional connectives from both the oral and the written modes. Our first research question is the following:

RQ1: Is competence with monofunctional connectives in teenage years predicted by their modality, i.e., whether they are used more frequently in oral or written language?

H1: Teenagers should use written connectives less accurately than oral ones, since they start to be exposed to written connectives mostly in secondary school, when they become more independent and proficient readers (Nippold, 2004, 2008).

This study will also shed light on the predictors of correct connective use related to some individual characteristics of teenagers, such as vocabulary level, exposure to print, and age. Our second research question can be summarized as follows:

RQ2: Is the use of various types of discourse connectives predicted by a broader lexicon size?

H2: We predict that general vocabulary knowledge contributes to connective use by teenagers, as these linguistic items represent a specific area of the lexicon (Crosson & Lesaux, 2013).

RQ3: Are teenagers who are more exposed to print also better at using different types of connectives?

H3: In line with previous studies on adults (Wetzel, 2020; Zufferey & Gygax, 2020a), it is likely that speakers who have a greater exposure to print will be better at using connectives, as it is mostly through the exposure to the written language that the biggest variety of connectives can be learnt.

RQ4: Does the competence to appropriately use connectives increase with age during teenage years?

H4: Regarding the role of age, we hypothesize that, in this age group, biological age will play a less prominent role than vocabulary size and exposure to print (Tskhovrebova *et al.*, 2022).

## Method

All materials, data, and code are available on the OSF repository ([https://osf.io/cbrsg/?view\\_only=9b50914daea04b6c9dabc91083520dcf](https://osf.io/cbrsg/?view_only=9b50914daea04b6c9dabc91083520dcf)). According to the regulations of the foundation, which granted funding on this study, we did not have to receive institutional ethics committee approval. However, all schools and adult participants gave their informed consent to participate in the present study.

## Participants

The participants of this study were 154 French-speaking students aged 12–19<sup>1</sup> ( $M = 14.43$ ,  $SD = 1.8$ , 80 females). All the participants were typically developing native speakers, as confirmed by their language teachers. The experiment was held in nine classes of four schools in the French-speaking part of Switzerland. Pupils came from the 9<sup>th</sup> ( $n = 53$ ,  $M_{\text{age}} = 12.57$ ,  $SD = 0.54$ ), 10<sup>th</sup> ( $n = 26$ ,  $M_{\text{age}} = 13.73$ ,  $SD = 0.78$ ), and 11<sup>th</sup> ( $n = 14$ ,  $M_{\text{age}} = 14.79$ ,  $SD = 0.80$ ) years of secondary school, and the first year of high school ( $n = 61$ ,  $M_{\text{age}} = 16.26$ ,  $SD = 0.95$ ). All schools gave their informed consent for participation in the study. We also tested a group of adults to determine the baseline of competence with connectives. For this purpose, we recruited 52 French speakers ( $M_{\text{age}} = 30.75$ ,  $SD = 11.07$ , range 19–58, 27 females) via the crowdsourcing platform Prolific© (Prolific, Oxford, UK, [www.prolific.co](http://www.prolific.co)). All participants showed at least 90% of good ratings in previous studies on the platform and gave their informed consent for participation in the study.



**Table 1.** Distribution of connectives per type of coherence relation and mode with their mean subjective orality rate ( $M_{OR}$ ) and frequency (per million words) in oral (Freq OR) and written (Freq WR) corpora

| Relation    | Mode    | Connective          | Translation in English    | $M_{OR}$ ( $SD$ ) | Freq OR | Freq WR |
|-------------|---------|---------------------|---------------------------|-------------------|---------|---------|
| Addition    | Oral    | <i>en plus</i>      | in addition (less formal) | 17.57 (2.97)      | 501.5   | 57.09   |
|             | Written | <i>en outre</i>     | in addition (more formal) | 3.63 (3.80)       | 1.75    | 91.29   |
| Cause       | Oral    | <i>parce que</i>    | because (less formal)     | 18.04 (2.71)      | 4086.75 | 262.42  |
|             | Written | <i>car</i>          | because (more formal)     | 11.75 (5.44)      | 68      | 552.85  |
| Concession  | Oral    | <i>même si</i>      | even if                   | 16.08 (3.06)      | 165.5   | 101.76  |
|             | Written | <i>néanmoins</i>    | nevertheless              | 5.43 (4.08)       | 7.25    | 86.73   |
| Consequence | Oral    | <i>donc</i>         | so                        | 17.27 (3.04)      | 5913    | 723.83  |
|             | Written | <i>ainsi</i>        | therefore                 | 8.14 (5.15)       | 64.5    | 292.72  |
| Contrast    | Oral    | <i>par contre</i>   | instead                   | 16.25 (3.29)      | 251.75  | 15.47   |
|             | Written | <i>en revanche</i>  | conversely                | 9.10 (4.31)       | 6.5     | 46.92   |
| Temporality | Oral    | <i>dès que</i>      | as soon as (less formal)  | 14.61 (3.97)      | 87      | 52.23   |
|             | Written | <i>aussitôt que</i> | as soon as (more formal)  | 7.96 (5.03)       | 0.5     | 7.44    |

## Materials

### Sentence cloze test

**Selection of connectives.** In this task, we tested six types of coherence relations (addition, contrast, temporality, consequence, cause, and concession), most common in corpus data (see, e.g., Prasad et al., 2008), each represented by one connective more typical for written language and another one more typical for oral speech (see Table 1 for the distribution of different types of connectives). Our distinction between oral and written connectives was based on a corpus analysis of connective frequencies in corpora of oral and written language and was also confirmed by the native speakers' judgments.

To calculate the connectives' frequency per million words in oral speech, we chose the oral subcorpus of French *Orféo* (Benzitoun et al., 2016), as it includes 4 million words and contains speech from a wide variety of genres, such as everyday conversation and public speech. The frequency of connectives in writing was calculated based on three corpora from different genres, including journalistic (Le Monde corpus), argumentative (the French part of the Europarl corpus, Koehn, 2005), and literary texts (the Frantext corpus, ATILF, 1998–2022). We first calculated the connective frequencies per million words separately for each corpus and then calculated the mean frequency for each connective. Connectives that were more frequently used in oral than in written corpora were categorized as *oral*, and connectives that were more frequent in written than in oral corpora were categorized as *written*.

To verify the outcomes of the corpus analysis, we asked a group of adults to judge to what extent each of the connectives chosen for the task was common in oral conversation in informal contexts (such as family dinner or a conversation with friends), on a scale from 0 to 20. The answer 0 meant that a connective is never

used in informal oral speech and 20 that it is used very frequently in this context. For every coherence relation, the connective with a higher score was labeled as *oral* and that with a lower score as *written*. The judgment task was performed online by native French speakers ( $N = 102$ ). None of them participated in the main experiment. The distinction between the connectives typically used in oral and written modes, as determined by native speakers' judgments, matched the outcome of the corpus analysis of connectives' frequencies.

*Structure of the test.* We asked participants to fill in gaps between two sentences with an appropriate connective. The gap was always in the initial position of the second sentence. The test included 60 items in total, 10 items per coherence relation, five of which targeted a written connective and other five oral ones. In the task, participants always had a choice between four options randomly selected out of six connectives tested in each mode. Consequently, if the expected answer was a written connective, the proposed options also belonged to this mode, and vice versa for the oral connectives (see examples for the relation of causality (7)–(8)). This allowed us to test the two modes separately and prevented participants from always choosing the oral connectives which are more common in everyday speech. For each experimental item, there was only one possible answer. The final score was calculated as the proportion of correct answers per connective.

(7) The correct answer: written connective *car* "for"

Le vendeur était très content de sa semaine // \_\_\_\_\_ // il avait réalisé d'excellentes ventes.

"The shop assistant was very happy with his week // \_\_\_\_\_ // he had made excellent sales."

Answer options: (a) *néanmoins* "however"; (b) *en revanche* "in contrast"; (c) *car* "for"; (d) *ainsi* "therefore"

(8) The correct answer: oral connective *parce que* "because"

Sarah était scandalisée // \_\_\_\_\_ // elle s'était fait licencier après vingt ans de bons et loyaux services.

"Sarah was outraged // \_\_\_\_\_ // she had been fired after twenty years of loyal service."

Answer options: (a) *donc* "so"; (b) *en plus* "also"; (c) *parce que* "because"; (d) *même si* "even though"

#### *Vocabulary level test*

To assess the vocabulary level of participants, we created a French version of a vocabulary size test based on Nation and Beglar (2007). The participants were asked to read a definition of a word and choose one of the six words that was the best match for the definition. The test included four categories of words, based on frequency lists from the French corpus *Lexique 3.83* (New *et al.*, 2001). Each category consisted of 30 items, which were selected from the first, second, third, and fourth 5000-word families. Moreover, each word category included different parts of speech, namely 18 nouns, 6 verbs, and 6 adjectives. Importantly, the foils also

belonged to the same frequency level as the target words. The word frequencies therefore decreased from the first to the fourth category, and the participants completed the task in the order of increasing frequencies. Vocabulary scores used the proportion of correct answers per participant. The reliability of the vocabulary test, as measured by Cronbach's alpha, was high both for teenagers and adults. For teenagers, it was of .96 (95% CI [.93–.97]<sup>2</sup>), and for adults, .91 (95% CI [.85–.93]).

#### *Author recognition test*

We developed a new version of the ART to assess teenagers' degree of exposure to print, as this test is not only sensitive to cultural differences (e.g., Stainthorp, 1997) but also to the age of participants (e.g., Cunningham & Stanovich, 1990). Our version of the ART (*ART-F-CL*) was based on the names of authors who are considered to be classics according to the listings of three big national chains of bookstores in Switzerland. The list included 40 author names and 40 names of unknown people, which were randomly mixed. The participants had to select only those names that they knew to be authors. The instruction mentioned that some of the names were not authors, and that one point would be removed if the participants checked the wrong name. For each correct answer, participants were given 1 point, and for each wrong one –1. We computed the general score summing up the points for correct and incorrect answers. The maximum possible score was 40 and the minimum –40.

For the group of adults, we used a different version of the ART (*ART-F*), developed for French by Zufferey and Gygas (2020a). It replicated the design of the original English ART (Stanovich & West, 1989) and was based on the names of best-selling and prize-winning authors (see <https://osf.io/yxj8q/> for the full task). The number of items and the calculation of the final score were the same as for the teenage version of the task described before. The reliability of the two ART tests was quite high, as indicated by their Cronbach's alphas (*ART-F-CL*: .88 [.85–.91]; *ART-F*: .92 [.86–.94]).

In addition to the ART, all the participants were asked to give a subjective evaluation of their exposure to print. In a separate question, they were asked to estimate how regularly they read on a scale ranging from 0 = *never* to 10 = *every day*.

#### **Procedure**

All the tasks were administered online via a weblink. The link was distributed directly among the teachers of the participating classes in the case of teenagers, and via the Prolific platform (<https://www.prolific.co>) in the case of adults. The order of the tasks was always the same. The participants started with the connective choice task and then proceeded to the ART and finished with the vocabulary test. Once the participants gave an answer and proceeded to the next question, they could not go back and correct their initial response. There was no time limit for the task, but the participants had to finish it in one session. The teenagers spent on average 1 hr on all the tasks, and it took adults approximately 40 min to complete the test battery.

## Analysis

We analyzed the correctness of responses in the cloze test, using a generalized mixed-effects logistic regression model and an automated backward selection using the statistical software R (R Core Team, 2012). Accuracy of responses ( $1 = \text{right}$ ,  $0 = \text{wrong}$ ) in the cloze task was the dependent variable in this analysis. We used an automatic backward selection because this way we could include all the tested predictors in the initial model and then automatically eliminate the nonsignificant ones. A forward selection procedure was deemed less appropriate in this case because mastery of connectives in teenage years has barely been investigated and we have no theoretical reasons for adding predictors to the model in a particular order. The initial full model was built with the *glmer* function of the *lme4* package (Bates et al., 2015) and included vocabulary size, exposure to print, subjective evaluation of exposure to print, age, and connective mode as predictors of performance on the connective task. All the variables of individual difference were centered. Since ART-CL was highly correlated with the vocabulary score ( $\rho = .50$  [.37, .61],  $p < .001$ ) and age ( $\rho = .49$  [.36, .61],  $p < .001$ ), age and vocabulary score were residualized by the ART-CL score by means of the *umx\_residualize* function of the *umx* package (Bates, 2021) to avoid multicollinearity in the statistical model.

Next, we conducted an automated selection of relevant predictors with *drop1* function of the *stats* package (R Core Team, 2012), deleting the fixed effects with the  $p$  values higher than .05. The outcome of the final reduced model was then returned with the *summary* function of the *lmerTest* package (Kuznetsova et al., 2017). The statistical significance level was set at 5% throughout the paper. Following the procedure by Schreiber-Gregory (2018), we controlled that the assumptions of logistic regressions were met (i.e., appropriate outcome structure, absence of multicollinearity, linearity of independent variables and log odds, and an appropriate sample size). Since our experiment had a repeated measures design (in that the same participants completed multiple test items, and the same test items were taken by multiple participants), the assumption of observation independence was not met. We however accounted for it by adding the random effects as intercepts for items and participants in our mixed-effects models.

Finally, we performed a random forest analysis (Strobl et al., 2009) based on the predictors included in the final reduced model in order to compare the impact of each relevant predictor variable on the dependent one (i.e., correctness of responses in the cloze task). The advantage of this method is that it does not have assumptions about the distribution of data and can make predictions even about highly correlated variables. Moreover, it is highly reliable, as variable importance is calculated based on a multitude of classification, or regression, trees (Strobl et al., 2009).

## Results

### *Descriptive statistics for the background measures*

As is evident from the descriptive statistics in Table 2, across all three measures, teenagers on average had lower scores than adults. The vocabulary level of the teenagers was about 22% lower than that of adults. The scores in ART were 2.2 points

**Table 2.** Descriptive statistics for background measures, by group

|                                     | <b>M (SD)</b> | Observed range | Possible range |
|-------------------------------------|---------------|----------------|----------------|
| <b>Vocabulary size</b>              |               |                |                |
| Teenagers                           | .71 (.15)     | .15 to .95     | 0 to 1         |
| Adults                              | .91 (.07)     | .72 to .99     |                |
| <b>Author recognition test*</b>     |               |                |                |
| Teenagers                           | 6.55 (5.93)   | -11 to 28      | -40 to 40      |
| Adults                              | 8.75 (7.63)   | -1 to 33       |                |
| <b>Subjective exposure to print</b> |               |                |                |
| Teenagers                           | 4.91 (3.10)   | 1 to 10        | 0 to 10        |
| Adults                              | 6.30 (2.28)   | 1 to 10        |                |

\*A different version of the ART was used for teenagers and adults.

lower in teenagers than in adults. Finally, the subjective evaluation of exposure to print was about 1 point lower for teenagers than for adults.

### **Results for the connective test**

The teenagers performed on average quite well in the connective insertion task. Even though they did not reach the adult level of competence for all connectives, their scores were close to those of adults, especially for the connectives of cause, contrast, and time (see Figure 1). Teenagers had the lowest scores for the connective *en outre* with .58 accuracy, followed by *en plus* and *ainsi* with approximately .75 accuracy, then concessive *même si* (.80) and *néanmoins* (.79), and finally, all the remaining connectives with more than .85 of correct responses.

The final reduced model based on the step-down selection of predictors included vocabulary level, exposure to print, age, and mode (Table 3). Adding interaction between vocabulary test and exposure to print did not improve the model fit ( $\chi^2(1) = 0.51, p = .475$ ). The estimates of the final model revealed that vocabulary level, exposure to print, and age were the most important predictors for the performance in the cloze test. Moreover, we observed that connectives mostly bound to writing tend to be slightly more challenging for teenagers than the ones used in speech, as demonstrated by an estimated decrease of  $0.49 \pm 0.25$  SD. The only discourse relation where this trend was not attested is causality, for which the written connective *car* had a very similar score (.93) to that of the spoken connective *parce que* (.91).

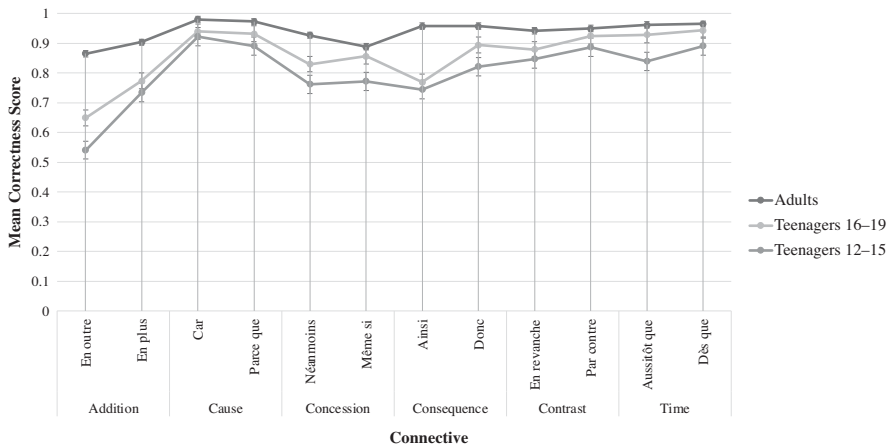
The overall prediction accuracy of the random forest analysis was 86%. This analysis supported the mixed logistic regression analysis and showed that vocabulary level had the most impact on the performance with connectives, followed by the score in ART-F-CL, age, and to a lesser extent mode (see Figure 2 for the visualization of the hierarchy of variable importance).

**Table 3.** Output of the full model and the final reduced model

| Variable                      | Full model |      |       |          | Final reduced model |      |       |          |
|-------------------------------|------------|------|-------|----------|---------------------|------|-------|----------|
|                               | Estimate   | SE   | z     | Pr(> z ) | Estimate            | SE   | z     | Pr(> z ) |
| (Intercept)                   | -0.69      | 0.55 | -1.26 | 0.209    | -0.47               | 0.53 | -0.89 | 0.376    |
| Vocabulary test**             | 5.23       | 0.78 | 6.74  | <0.001   | 5.37                | 0.77 | 6.93  | <0.001   |
| ART-F-CL*                     | 0.94       | 0.17 | 5.44  | <0.001   | 0.96                | 0.17 | 5.51  | <0.001   |
| Age*                          | 1.86       | 0.64 | 2.92  | 0.004    | 1.63                | 0.62 | 2.63  | 0.008    |
| Subjective exposure to print* | 0.17       | 0.12 | 1.40  | 0.161    |                     |      |       |          |
| Written mode                  | -0.49      | 0.25 | -1.96 | 0.050    | -0.49               | 0.25 | -1.96 | 0.050    |

\*Centered values.

\*\* Centered and residualized values.



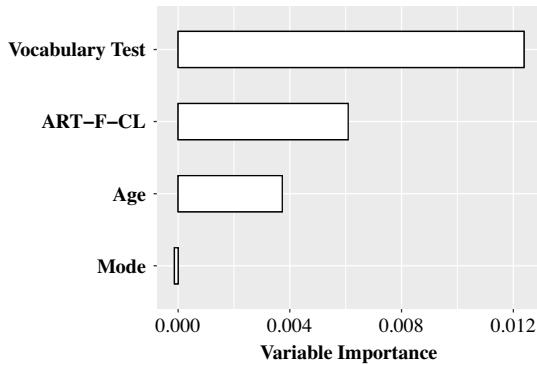
**Figure 1.** Distribution of Mean Scores per Connective in Sentence Cloze Task among French-Speaking Teenagers and Adults.

*Note.* We included age as continuous variable in our statistical model because it is a more robust statistical approach compared to splitting the sample into (arbitrarily determined) age groups (Goldstein, 1979; Mirman, 2014). However, on the graph, we presented the results for teenagers in two age groups, namely 12–15 (secondary school) and 16–19 (high school).

## Discussion

### **Competence with monofunctional connectives**

The goal of the current study was to provide new evidence on the level of competence with discourse connectives by students during teenage years. More precisely, this research aimed to explore student- and connective-related factors that could explain variability in the level of connective mastery in French-speaking teenagers. Our results show that, on average, teenagers have a good command of all the monofunctional connectives included in our experiment. This finding indicates that, once the challenge of polyfunctionality is removed, teenagers can successfully perform on the cloze task including connectives of different types in both modes.



**Figure 2.** The Impact of Each Predictor Variable on the Dependent Variable According to the Random Forest Analysis.

Teenagers had the most difficulty with the additive connective *en outre* “in addition (more formal).” The fact that another additive connective *en plus* “in addition (less formal)” was mastered rather well suggests that the difficulty with *en outre* cannot be explained by the complexity of the additive coherence relation but rather by particular features of this connective. As shown by the preliminary studies assessing the degree of orality of the tested connectives, *en outre* received the lowest orality scores (Table 1). This means that native speakers perceive it as extremely uncommon in spoken language and probably have less clear intuitions about its usage, as they may be not exposed enough to the written contexts where this connective is used.

We have also systematically assessed whether the mode in which connectives are typically used (written vs. spoken) has an impact on their general mastery. Our results indicate that the connectives bound to writing were slightly more challenging for teenagers than ones used in speech, even though the differences between them were rather small. It would be legitimate to assume that the difference between the mastery of oral and written connectives may result from the overall higher frequency of oral connectives. However, this reasoning does not quite apply to all the connectives tested in this study. For instance, written connectives *en outre*, *car*, and *en revanche*, which are infrequent in oral corpora, are more frequent than their oral counterparts in written corpora (see Table 1 for frequencies). Therefore, at least for these connectives, the difference in scores may be explained by the fact that teenagers had not been exposed enough to the written contexts in which these connectives are used more frequently than in spoken language. Furthermore, these results may also indicate that performance with connectives from the written mode does not come naturally to the same extent as performance with those used in spoken language. Getting access to written connectives requires more effort, as exposure to them comes only through reading. Hence, school curricula should devote more time to teach this type of connectives as part of written language competence.

Although teenagers performed well in the connective insertion task, their performance was still inferior to that of adults and showed large individual variability. This finding is in line with previous research on language development in older children

(Berman, 2004; Nippold, 2008) suggesting that adult-level language proficiency is acquired far beyond puberty and that proficiency with connectives continues to develop even after the high-school years, as high-schoolers are still not adult like in their ability to use appropriate connectives. It is likely that if we used a more challenging task in which teenagers had to insert connectives within a short text, the difference in the performance between adults and teenagers would be even higher, as it was found in previous research (Tskhovrebova *et al.*, 2022). The attested difference between teenagers and adults also suggests that the ability to use connectives does not come solely with cognitive maturation, but rather it crucially hinges on a more extensive general linguistic experience that is gained throughout the lifespan.

### ***Student-level predictors of mastery of connectives***

As evidenced by all the statistical results, it is teenagers' vocabulary level rather than chronological age that appears to be the strongest predictor of the appropriate usage of connectives in French. This result suggests that, even though connectives may differ from other vocabulary items, as they guide speakers in the interpretation of discourse rather than express a particular concept (e.g., Wilson & Sperber, 1993), a higher vocabulary level still significantly contributes to a better usage of monofunctional connectives already during teenage years. We believe that this finding does not necessarily undermine the idea that connectives are processing instructions. It rather confirms their intermediate nature as specific lexical items (declarative knowledge) expressing procedural meaning (procedural knowledge). It is thus plausible to assume that the acquisition of connectives is supported by both declarative and procedural memory systems. Future research should empirically test this assumption by relating competence with connectives to procedural and declarative learning ability.

On the other hand, participants with a greater exposure to print may benefit from other linguistic skills (e.g., reading or sentence-processing ability), which could compensate for the lack of vocabulary, when filling in the connective task. As a matter of fact, the degree of exposure to written language, as measured by the ART, appears to be another important factor accounting for variations in the competence with connectives. This finding thus supports previous research on this matter in adults (Zufferey & Gygas, 2020a) and extends its validity on younger participants. The fact that exposure to print and the mastery of connectives are related suggests that long-term reading habits, as revealed by the ART (Scholman *et al.*, 2020), may help to acquire linguistic experience that is necessary for an accurate use of connectives in discourse. By linguistic experience, we mean a complex set of linguistic components, such as vocabulary knowledge (see, e.g., Stanovich *et al.*, 1995), reading comprehension (see, e.g., Spear-Swerling *et al.*, 2010), and competence with meta-cognitive analysis of texts (McBride-Chang & Chang, 1995). The performance with the ART is related to all of them, but it cannot be reduced to any of them. Moreover, the knowledge of authors' names may shed light on the general cultural capital and socioeconomic status (SES) of participants, and the latter was found to be related to the competence with connectives already in primary school years (Volodina &



Weinert, 2020). Therefore, future research should examine in more detail the relation between SES, exposure to print, and the usage of connectives.

In contrast to previously developed ARTs, which privileged popular out-of-school readings among a specific age group of speakers from a specific region (e.g., Allen et al., 1992; Cunningham & Stanovich, 1991; Spear-Swerling et al., 2010), the novel ART-CL, based on classical authors, is less geographically anchored and more polyvalent, as a list of classic literature can be easily found in school curriculum guidelines and catalogues of big bookstore chains. As regards the subjective evaluation of exposure to print, it was found to be less adequate than the ART tests, as it did not account at all for the individual variation of the connective use. This result may stem from the fact that such a measure of exposure to print may be subject to the production of socially desirable answers and guessing (e.g., Chateau & Jared, 2000; Echols et al., 1996). As a result, it might indicate attitude toward reading rather than degree of exposure to print as such (e.g., Allen et al., 1992). For future studies, we would therefore recommend to rely on more robust measures of exposure to print, such as the ART used in the present work.

Age was also an important predictor for the correct usage of connectives in the cloze test, according to both methods of statistical analysis. This indicates that cognitive maturation and linguistic experience, increasing during teenage years and early adulthood, are crucial for the mastery of connectives. Nevertheless, this factor proved to be less important than vocabulary level and degree of exposure to print, as measured by ART-CL. This result is in stark contrast to prior research findings demonstrating that age is a strong predictor of connective acquisition in young children (Blything et al., 2015; Cain & Nash, 2011; Pyykkönen & Järvikivi, 2012; Volodina & Weinert, 2020). This finding demonstrates that later language development, such as the developing ability to use a broad range of connectives, is qualitatively different from early language acquisition and happens in a slower and qualitatively different way than in early childhood (Nippold, 1993). This means that competence with connectives does not come naturally (through cognitive maturation) but rather requires extensive input through reading. This result also supports the Ullman's (2001) DP model of language, suggesting that connectives may indeed be part of procedural knowledge, as age turns out to be less important for the mastery of connectives during teenage years when procedural learning slows down.

### Limitations and future directions

It is possible that the scores in the connective insertion task were quite high because of the task design. In real life, to express certain coherence relations speakers have to retrieve relevant connectives from their mental lexicon and choose an appropriate connective out of a multitude of existing connectives, varying in frequency, poly-functionality, and bearing different syntactic constraints. In contrast, our task provided only four answer options for each pair of sentences, this way considerably limiting the choice of connectives and simplifying the challenge. In other words, the findings presented in this paper do not necessarily mean that teenagers would perform well with all the connectives in real-life contexts. Rather, the current results suggest that teenagers manage to match a monofunctional connective with a

coherence relation in the context of a restricted number of connectives. Future research will need to provide more evidence on the acquisition of potentially more challenging connectives, such as those that have very low frequency and those that can be used to express multiple coherence relations.

Regarding the factor of modality, according to random forests and mixed logistic regression analysis, it was less important than all the other predictors, except for subjective exposure to print. We believe that the effect of mode may have been diminished due to our experimental design. It is possible that connectives mostly used in writing are not difficult *per se*, but that it is more challenging to use them in formal written contexts due to the overall lower frequency of certain written connectives and to the complexity of their linguistic form and of the topics that are covered in written texts from various genres. This complexity, however, was completely neutralized in our experiment, as we tested connective use in isolated pairs of sentences within a strictly monofunctional and simple context. Future research should strive to include more ecological contexts and include also online processing measures in order to complement these results.

Finally, the fact that certain connectives, such as *en outre*, are still challenging for teenagers suggests that not all connectives are fully acquired by this group of speakers. Future work should therefore focus more on infrequent connectives that have even lower orality ratings, as well as on the less frequent functions of polyfunctional connectives. Another important dimension for future work will be to assess how connectives are taught at school, and how teaching methods can be made more efficient.

## Conclusion

Knowledge of connectives is crucial for reading comprehension and overall academic performance. Our results demonstrate that teenagers have not yet attained adult-like mastery of connectives and there is significant individual variability in connective knowledge. In our study, performance on the connective task was strongly predicted by vocabulary level and degree of exposure to print, and only to a lesser extent by student age and linguistic mode of the connective. This finding stresses the need to enhance students' exposure to print during teenage years. More frequent exposure to written texts will provide students with actual examples of connectives in use and allow them to practice using connectives as processing instructions for building a mental model of the coherence relations in a text. Exposure to print also plays a key role in enlarging student vocabularies and, as shown by our results, vocabulary size is another strong predictor of competence with connectives, on top of students' exposure to print.

Our study contributes to the research on linguistic development during teenage years. It highlights that, by the end of high school, French-speaking teenagers still do not attain an adult level of competence with monofunctional connectives from written and oral modes. This means that linguistic proficiency continues to develop far beyond puberty and during late teenage years. The focus of language acquisition research has been mainly placed on language development in pre-school and early school years. Our findings highlight the need to study acquisition processes in later

childhood and adolescence. Identifying linguistic domains that are still problematic in adolescence will help develop teaching materials that could enhance students' ability to understand written texts and oral explanations by the teacher and thereby contribute to general academic achievement.

**Replication package.** Data, materials, and code are openly available at the project's Open Science Framework page ([https://osf.io/cbrsg/?view\\_only=9b50914daea04b6c9dabc91083520dcf](https://osf.io/cbrsg/?view_only=9b50914daea04b6c9dabc91083520dcf)).

**Financial support.** This work was funded by Swiss National Science Foundation Grant 100012\_184882.

**Conflict of interests.** The authors declare none.

## Notes

1 We presented the results of two participants who were 19 years old together with the group categorized as teenagers because they were recruited together with other students of the first year of high school and followed the same curriculum in French as their classmates. We did not consider it appropriate to present the results of these two participants, who were not more advanced than their younger classmates, together with the group categorized as adults, who had already finished their school studies and were recruited in a different context, namely via a crowdsourcing platform.

2 In square brackets, we reported 95% confidence intervals.

## References

- Akbaşlı, S., Şahin, M., & Yaykiran, Z. (2016). The effect of reading comprehension on the performance in science and mathematics. *Journal of Education and Practice*, 7(16), 108–121. <https://eric.ed.gov/?id=EJ1108657>
- Allen, L., Cielewski, J., & Stanovich, K. E. (1992). Multiple indicators of children's reading habits and attitudes: Construct validity and cognitive correlates. *Journal of Educational Psychology*, 84(4), 489–503. <https://doi.org/10.1037/0022-0663.84.4.489>
- Asr, F. T., & Demberg, V. (2020). Interpretation of discourse connectives is probabilistic: Evidence from the study of 'but' and 'although'. *Discourse Processes*, 57(4), 376–399. <https://doi.org/10.1080/0163853X.2019.1700760>
- ATILF. (1998–2022). *Base textuelle Frantext (En ligne) [Data set]*. ATILF-CNRS & Université de Lorraine. <https://www.frantext.fr/>
- Barr, C. D., Uccelli, P., & Phillips Galloway, E. (2019). Specifying the academic language skills that support text understanding in the middle grades: The design and validation of the core academic language skills construct and instrument. *Language Learning*, 69, 978–1021. <https://doi.org/10.1111/lang.12365>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Bates, T. (2021). *umx: A helper package for Structural Equation Modeling in OpenMx*. Edinburgh, UK: University of Edinburgh. <https://doi.org/10.5281/zenodo.10937>
- Beek ter, M. (2020). *Supporting reading comprehension in history education: The use and usefulness of a digital learning environment*. University of Groningen. <https://doi.org/10.33612/diss.121518620>
- Benzitoun C., Debaisieux, J.-M., & Deulofeu, J. (2016). Le projet ORFÉO: un corpus d'étude pour le français contemporain. *Corpus*, 15. <https://doi.org/10.4000/corpus.2936>
- Berman, R. (2004). Between emergence and mastery: The long developmental route of language acquisition. In R. Berman (Ed.), *Language development across childhood and adolescence* (pp. 9–34). John Benjamins Publishing Company. <https://doi.org/10.1075/tilar.3>
- Berman, R. (2004). *Language development across childhood and adolescence*. John Benjamins Publishing Company. <https://doi.org/10.1075/tilar.3>
- Blakemore, D. (2002). *Relevance and linguistic meaning. The semantics and pragmatics of discourse markers*. Cambridge: Cambridge University Press.

- Blything, L. P., Davies, R., & Cain, K. (2015). Young children's comprehension of temporal relations in complex sentences: The influence of memory on performance. *Child Development*, *86*(6), 1922–1934. <https://doi.org/10.1111/cdev.12412>
- Britton, B. K. (1994). Understanding expository text. Building mental structures to induce insights. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 641–674). San Diego, CA: Academic Press.
- Cain, K., & Nash, H. M. (2011). The influence of connectives on young readers processing and comprehension of text. *Journal of Educational Psychology*, *103*(2), 429–441. <https://doi.org/10.1037/a0022824>
- Canestrelli, A. R., Mak, W. M., & Sanders, T. J. M. (2013). Causal connectives in discourse processing: How differences in subjectivity are reflected in eye movements. *Language and Cognitive Processes*, *28*(9), 1394–1413. <https://doi.org/10.1080/01690965.2012.685885>
- Chateau, D., & Jared, D. (2000). Exposure to print and word recognition processes. *Memory & Cognition*, *28*, 143–153. <https://doi.org/10.3758/BF03211582>
- Corkett, J. K., Parrila, R., & Hein, S. F. (2006). Learning and study strategies of university students who report a significant history of reading difficulties. *Developmental Disabilities Bulletin*, *34*, 57–79.
- Cribble, L., & Cuenca, M. (2017). Discourse markers in speech: Characteristics and challenges for annotation. *Dialogue and Discourse*, *8*(2), 149–166. <http://hdl.handle.net/1854/LU-8727154>
- Crosson, A., & Lesaux, N. (2013). Does knowledge of connectives play a unique role in the reading comprehension of English learners and English-only students? *Journal of Research in Reading*, *36*, 241–260. <https://doi.org/10.1111/j.1467-9817.2011.01501.x>
- Cunningham, A., & Stanovich, K. (1990). Assessing print exposure and orthographic processing skill in children: A quick measure of reading experience. *Journal of Educational Psychology*, *82*(4), 733–740. <https://doi.org/10.1037/0022-0663.82.4.733>
- Cunningham, A., & Stanovich, K. (1991). Tracking the unique effects of print exposure in children: Associations with vocabulary, general knowledge, and spelling. *Journal of Educational Psychology*, *83*(2), 264–274. <https://doi.org/10.1037/0022-0663.83.2.264>
- Degand, L., & Sanders, T. J. M. (2002). The impact of relational markers on expository text comprehension in L1 and L2. *Reading and Writing: An Interdisciplinary Journal*, *15*, 739–757. <https://doi.org/10.1023/A:1020932715838>
- Echols, L. D., West, R. F., Stanovich, K. E., & Zehr, K. S. (1996). Using children's literacy activities to predict growth in verbal cognitive skills: A longitudinal investigation. *Journal of Educational Psychology*, *88*, 296–304. <https://doi.org/10.1037/0022-0663.88.2.296>
- Evers-Vermeul, J., & Sanders, T. (2009). The emergence of Dutch connectives; how cumulative cognitive complexity explains the order of acquisition. *Journal of Child Language*, *36*(4), 829–854. <https://doi.org/10.1017/S0305000908009227>
- Fuentes, P. (1998). Reading comprehension in mathematics. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *72*, 81–88. <https://doi.org/10.1080/00098659809599602>
- Gernsbacher, M. A. (1997). Coherence cues mapping during comprehension. In J. Costermans & M. Fayol (Eds.), *Processing interclausal relationships. Studies in the production and comprehension of text* (pp. 3–21). Mahwah, NJ: Lawrence Erlbaum Associates.
- Goldstein, H. (1979). *The design and analysis of longitudinal studies: Their role in the measurement of change*. London: Academic Press.
- Halliday, M., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Imam, O. A., Mastura, M. A., Jamil, H., & Ismail, Z. (2014). Reading comprehension skills and performance in science among high school students in the Philippines. *Asia Pacific Journal of Educators and Education*, *29*, 81–94.
- Jordan, N. C., Kaplan, D., Olah, L.N., & Locuniak, M. N. (2006). Number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematics difficulties. *Child Development*, *77*(1), 153–175. <https://doi.org/10.1111/j.1467-8624.2006.00862.x>
- Kleijn, S., Pander Maat, H. L. W., & Sanders, T. J. M. (2019). Comprehension effects of connectives across texts, readers, and coherence relations. *Discourse Processes*, *56*(5–6), 447–464. <https://doi.org/10.1080/0163853X.2019.1605257>
- Koehn, P. (2005). Europarl: A parallel corpus for statistical machine translation. *Conference Proceedings: The Tenth Machine Translation Summit, Phuket, Thailand* (pp. 79–86). <https://homepages.inf.ed.ac.uk/pkoehn/publications/europarl-mtsummit05.pdf>

- Koehne, J., & Demberg, V. (2013). The time-course processing of discourse connectives. *Proceedings of the 35th Annual Meeting of the Cognitive Science Society (CogSci2013)*, 2760–2765.
- Korpershoek, H., Kuiper, H., & van der Werf, G. (2015). The relation between students' math and reading ability and their mathematics, physics, and chemistry examination grades in secondary education. *International Journal of Science and Mathematics Education*, *13*, 1013–1037. <https://doi.org/10.1007/s10763-014-9534-0>
- Kuznetsova, A., Bruun Brockhoff, P., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26.
- Linderholm, T., Everson, M. G., van den Broek, P. W., Mischini, M., Crittenden, A., & Samuels, J. (2000). Effects of causal text revisions on more- and less-skilled readers' comprehension of easy and difficult texts. *Cognition and Instruction*, *18*, 525–556. [https://doi.org/10.1207/S1532690XCI1804\\_4](https://doi.org/10.1207/S1532690XCI1804_4)
- McBride-Chang, C., & Chang, L. (1995). Memory, print exposure, and metacognition: Components of reading in Chinese children. *International Journal of Psychology*, *30*, 607–616. <https://doi.org/10.1080/00207599508246589>
- McClure, E., & Geva, E. (1983). The development of the cohesive use of adversative conjunctions in discourse. *Discourse Processes*, *6*, 411–432. <https://doi.org/10.1080/01638538309544575>
- Millis, K., & Just, M. (1994). The influence of connectives on sentence comprehension. *Journal of Memory and Language*, *33*(1), 128–147.
- Millis, K. K., Graesser, A. C., & Haberlandt, K. (1993). The impact of connectives on the memory for expository texts. *Applied Cognitive Psychology*, *7*(4), 317–339. <https://doi.org/10.1002/acp.2350070406>
- Mirman, D. (2014). *Growth curve analysis and visualization using R*. CRC Press.
- Nation, I. S. P., & Coxhead, A. (2021). *Measuring native-speaker vocabulary size*. John Benjamins.
- Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*(7), 9–13.
- New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE [A lexical database of contemporary French: LEXIQUE]. *L'Année Psychologique*, *101*, 447–462. <http://www.lexique.org>
- Nippold, M. (1993). Developmental markers in adolescent language: Syntax, semantics, and pragmatics. *Language, Speech, and Hearing Services in Schools*, *24*, 21–28. <https://doi.org/10.1044/0161-1461.2401.21>
- Nippold, M. (2004). Research on later language development international perspectives. In R. Berman (Ed.), *Language development across childhood and adolescence* (pp. 1–8). John Benjamins Publishing Company. <https://doi.org/10.1075/tilar.3>
- Nippold, M. (2008). *Later language development: school-age children, adolescents, and young adults* (3rd ed., 2nd printing). PRO-ED.
- Nippold, M., Schwartz, I., & Undlin, R. (1992). Use and understanding of adverbial conjunctions: A developmental study of adolescents and young adults. *Journal of Speech and Hearing Research*, *35*, 108–118.
- O'Reilly, T., & McNamara, D. S. (2007). The impact of science knowledge, reading skill, and reading strategy knowledge on more traditional “high-stakes” measures of high school students' science achievement. *American Educational Research Journal*, *44*(1), 161–196. <https://doi.org/10.3102/0002831206298171>
- Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction*, *19*, 228–242. <https://doi.org/10.1016/j.learninstruc.2008.04.003>
- Peterson, C. (1986). Semantic and pragmatic uses of 'but'. *Journal of Child Language*, *13*, 583–590. <https://doi.org/10.1017/S0305000900006905>
- Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., & Webber, B. (2008). The Penn Discourse Treebank 2.0. *Proceedings of the 6th International Conference of Language Resources and Evaluation (LREC 2008)*, Marrakech, Morocco. <https://www.seas.upenn.edu/~pdtb/papers/pdtb-lrec08.pdf>
- Pyykkönen, P., & Järvikivi, J. (2012). Children and situation models of multiple events. *Developmental Psychology*, *48*(2), 521–529. <https://doi.org/10.1037/a0025526>
- RAND Reading Study Group, & Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. RAND Corporation. <http://www.jstor.org/stable/10.7249/mr1465oeri>
- R Core Team (2012). *R: A language and environment for statistical computing*. [Computer software manual]. Vienna, Austria. <http://www.R-project.org/>

- Roze, C., Danlos, L., & Muller, P. (2012). LEXCONN: A French lexicon of discourse connectives. *Discourse*, 10, 1–15.
- Salihi, L., Aro, M., & Räsänen, P. (2018). Children with learning difficulties in mathematics: Relating mathematics skills and reading comprehension. *Issues in Educational Research*, 28, 1024–1038. <http://www.iier.org.au/iier28/salihi.pdf>
- Sanders, T., & Spooren, W. (2007). Discourse and text structure. In D. Geeraerts & H. Cuykens (Eds.), *Handbook of cognitive linguistics* (pp. 37–60). Oxford: Oxford University Press.
- Sanders, T. J. M., Spooren, W. P. M., & Noordman, L. G. M. (1992). Toward a taxonomy of coherence relations. *Discourse Processes*, 15(1), 1–35. <https://doi.org/10.1080/01638539209544800>
- Schleicher, A. (2019). *PISA 2018. Insights and interpretations*. Paris: OECD Publishing.
- Scholman, M. C. J., Demberg, V., & Sanders, T. J. M. (2020). Individual differences in expecting coherence relations: Exploring the variability in sensitivity to contextual signals in discourse. *Discourse Processes*, 57(10), 844–861. <https://doi.org/10.1080/0163853X.2020.1813492>
- Schreiber-Gregory, D. N. (2018). Ridge regression and multicollinearity: An in-depth review. *Model Assisted Statistics and Applications*, 13, 359–365. <https://doi.org/10.3233/MAS-180446>
- Snow, C. E., & Uccelli, P. (2009). The challenge of academic language. In D. R. Olson & N. Torrance (Eds.), *The Cambridge handbook of literacy* (pp. 112–133). Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/CBO9780511609664>
- Spear-Swerling, L., Brucker, P. O., & Alfano, M. P. (2010). Relationships between sixth-graders reading comprehension and two different measures of print exposure. *Reading and Writing: An Interdisciplinary Journal*, 23(1), 73–96. <https://doi.org/10.1007/s11145-008-9152-8>
- Stainthorp, R. (1997). A children's author recognition test: A useful tool in reading research. *Journal of Research in Reading*, 20(2), 148–158. <https://doi.org/10.1111/1467-9817.00027>
- Stanovich, K., & West, R. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, 24(4), 402–433. <https://doi.org/10.2307/747605>
- Stanovich, K., West, R., & Harrison, R. (1995). Knowledge growth and maintenance across the life span. The role of print exposure. *Developmental Psychology*, 31(5), 811–826. <https://doi.org/10.1037/0012-1649.31.5.811>
- Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, 14(4), 323–348. <https://doi.org/10.1037/a0016973>
- Traxler, M. J., Sanford, A. J., Aked, J. P., & Moxey, L. M. (1997). Processing causal and diagnostic statements in discourse. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 88–101. <https://doi.org/10.1037/0278-7393.23.1.88>
- Tskhovrebova, E., Zufferey, S., & Gygax, P. (2022). Individual variations in the mastery of discourse connectives from teenage years to adulthood. *Language Learning*, 72(2). <https://doi.org/10.1111/lang.12481>
- Ullman, M.T. (2001). The declarative/procedural model of Lexicon and Grammar. *Journal of Psycholinguistic Research*, 30, 37–69. <https://doi.org/10.1023/A:1005204207369>
- van Silfhout. (2014). Fun to read or easy to understand? Establishing effective text features for educational texts on the basis of processing and comprehension research [Doctoral dissertation, University of Utrecht]. Utrecht University Repository.
- van Silfhout, G., Evers-Vermeul, J., & Sanders, T. J. M. (2015). Connectives as processing signals: How students benefit in processing narrative and expository texts. *Discourse Processes*, 52(1), 47–76. <https://doi.org/10.1080/0163853X.2014.905237>
- Volodina, A., & Weinert, S. (2020). Comprehension of connectives: Development across primary school age and influencing factors. *Frontiers in Psychology* 11, 814. <https://doi.org/10.3389/fpsyg.2020.00814>
- Welie, C., Schoonen, R., Kuiken, F., & van den Bergh, H. (2017). Expository text comprehension in secondary school: For which readers does knowledge of connectives contribute the most? *Journal of Research in Reading*, 40, S42–S65. <https://doi.org/10.1111/1467-9817.12090>
- West, R., Stanovich, K., & Mitchell, H. (1993). Reading in the real world and its correlates. *International Reading Association*, 28, 35–50.
- Wetzel, M., Zufferey, S., & Gygax, P. (2020). Second language acquisition and the mastery of discourse connectives: assessing the factors that hinder L2-learners from mastering French connectives. *Languages* 5, 35. <https://doi.org/10.3390/languages5030035>

- Wilson, D.** (2011). The conceptual-procedural distinction: past, present and future. In V. Escandell-Vidal, M. Leonetti & A. Ahern (Eds.), *Procedural meaning: Problems and perspectives* (pp. 3–31). Bingley: Emerald Group Publishing.
- Wilson, D., & Sperber, D.** (1993). Linguistic form and relevance. *Lingua*, **90**(1–2), 1–25.
- Zufferey, S., & Gygax, P.** (2020a). “Roger broke his tooth. However, he went to the dentist”: Why some readers struggle to evaluate wrong (and right) uses of connectives. *Discourse Processes*, **57**(2), 184–200. <https://doi.org/10.1080/0163853X.2019.1607446>
- Zufferey, S., & Gygax, P.** (2020b). Do teenagers know how to use connectives from the written mode? *Lingua*, **234**(102779), 1–12. <https://doi.org/10.1016/j.lingua.2019.102779>
- Zufferey S., Mak W., Degand L., & Sanders T.** (2015). Advanced learners’ comprehension of discourse connectives: The role of L1 transfer across on-line and off-line tasks. *Second Language Research*, **31**(3), 389–411.

---

**Cite this article:** Tskhovrebova, E., Zufferey, S., and Tribushinina, E. (2022). French-speaking teenagers’ mastery of connectives: the role of vocabulary size and exposure to print. *Applied Psycholinguistics* **43**, 1141–1163. <https://doi.org/10.1017/S0142716422000303>