Discourse Processes

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hdsp20

Vocabulary size and exposure to print predict mastery of connectives in teenage years

Ekaterina Tskhovrebova, Sandrine Zufferey & Elena Tribushinina

To cite this article: Ekaterina Tskhovrebova, Sandrine Zufferey & Elena Tribushinina (2023) Vocabulary size and exposure to print predict mastery of connectives in teenage years, Discourse Processes, 60:10, 722-740, DOI: 10.1080/0163853X.2023.2266963

To link to this article: https://doi.org/10.1080/0163853X.2023.2266963

Published online: 30 Dec 2023.

DISCOURSE PROCESSES
2023, VOL. 60, NO. 10, 722–740
https://doi.org/10.1080/0163853X.2023.2266963

Routledge
Taylor & Francis Group

Vocabulary size and exposure to print predict mastery of connectives in teenage years

Ekaterina Tskhovrebova [a], Sandrine Zufferey [a], and Elena Tribushinina [b]

[a]Department of French Language and Literature, University of Bern; [b]Department of Languages, Literature and Communication, Utrecht University

ABSTRACT

Connectives such as *because* and *but* are crucial for signaling coherence relations in discourse. They contribute to a better reading comprehension and, thus, academic performance. The aim of this article is to contribute to our understanding of connective development during teenage years by studying individual differences in the performance of native Russian-speaking teenagers (N = 107, M_{age} = 13.93, range: 11 to 17) in a connective-cloze task. The tested connectives marked six coherence relations and were used either predominantly in speech or in the written language. In addition, we examined whether students' performance with the connectives was modulated by their general linguistic experience, as assessed by a vocabulary test and degree of exposure to print. Our results reveal that interpersonal differences in lexicon size and level of exposure to print were the strongest predictors of appropriate usage of connectives, whereas differences in age, connective mode, and polyfunctionality played a lesser role. This finding may indicate that, starting from age 11, biological age and intrinsic properties of connectives matter less for their mastery than general linguistic experience, as measured by vocabulary level and exposure to print.

Introduction

Reading comprehension is an essential competence for academic success, as most teaching relies on transmitting knowledge through written texts. It is well-documented that poor reading comprehension skills may hinder success in various academic fields (see, e.g., Baştuğ, 2014; Salihu et al., 2018; ter Beek, 2020). For instance, to succeed in an examination in mathematics or social science, a student generally should be able not only to show their knowledge of the subject but also to understand the description of the task to fulfill it correctly. Task comprehension can be facilitated through the use of discourse connectives. These linguistic devices express underlying coherence relations, such as condition (1) or cause (2), and thus contribute to a better understanding of the links between parts of texts and of the whole text in general (e.g., Degand & Sanders, 2002; Millis et al., 1993).

(1) Mathematics:
Какова скорость автомобиля в километрах в час, если спидометр показывает 65 миль в час?
"What is the speed of a car in kilometers per hour if the speedometer shows 65 miles per hour?"

CONTACT Ekaterina Tskhovrebova ekaterina.tckhovrebova@unibe.ch Institut de Langue et de Littérature Françaises, University of Bern, Länggassstrasse 49, Bern 3012, Switzerland

(2) Social science:

Каждый пятый молодой человек не интересуется политикой, так как считает, что это не его дело.

"One in five young people are not interested in politics because they think it is none of their business."

Appropriate use of and understanding various types of connectives constitutes an integral part of advanced academic skills (see, e.g., Barr et al., 2019) that is mostly acquired during secondary and high school years (M. A. Nippold, 2004; Nippold, 2008). However, few studies targeting a handful of languages have examined connective competence during this period (see, e.g., for Dutch: Kleijn et al., 2019; for English: Nippold et al., 1992; for French: Tskhovrebova, Zufferey, & Gygax, 2022; Tskhovrebova et al., 2022a; Zufferey & Gygax, 2020b), which is fundamental for the development of adult-like linguistic mastery. Therefore, it is crucial to broaden the research on connectives in teenage years and to extend it to other languages. To this end, the present study focused on the use of connectives by Russian-speaking teenagers.

Tskhovrebova et al. (2022a) found that French-speaking teenagers aged from 12 to 19 had a good command of 12 monofunctional connectives, expressing the relations of addition, concession, contrast, temporality, cause, and consequence, typically used in oral or written language. They also demonstrated that this command was predicted by teenagers' general vocabulary size and their level of exposure to print, rather than by their age. This finding is important, as it shows that language development during the teenage years, involving the developing ability to use a wide variety of connectives, appears to be qualitatively different from early stages of language acquisition (Nippold, 1993). In fact, in the childhood years, age constitutes one of the major predictors of connective acquisition (see, e.g., Blything et al., 2015).

The study by Tskhovrebova et al. (2022a) focused on the use of monofunctional connectives to overcome the design bias related to the alternative functions of polyfunctional connectives, as it was, for instance, the case in the study of Tskhovrebova, Zufferey, and Gygax (2022). More precisely, the authors of the latter study examined the use of two monofunctional (*en outre* "moreover," *toutefois* "however") and two polyfunctional (*aussi* "therefore," *en effet* "for") connectives by French-speaking teenagers. In the error analysis, the authors observed that teenagers tended to mistakenly use the connective *aussi* "therefore" instead of the connective *en outre* "moreover." This is likely because *aussi* is polyfunctional and can signal not only the consequence relation but also the additive relation. However, *aussi* can be used with an additive meaning exclusively in sentence-medial or final position (Roze et al., 2012). This suggests that teenagers did not know the consequence function of *aussi* and used it as an additive connective in the sentence-initial position despite important syntactic constraints.

To avoid a similar design bias stemming from the alternative functions of polyfunctional connectives, Tskhovrebova et al. (2022a) examined only the performance with monofunctional French connectives. However, it is not completely clear whether teenagers' performance on the cloze test was quite high because monofunctional connectives are generally easier than polyfunctional ones or because of the removed design bias. Therefore, in the present study, we adopt a different strategy to solve this issue, without completely excluding polyfunctional connectives from the examination. We assessed the use of mono- and polyfunctional Russian connectives that are typically used in the written and oral modes and signal the same type of coherence relations as the French connectives studied by Tskhovrebova et al. (2022a). Yet, to address the design bias associated with the polyfunctionality of certain connectives and to avoid the possibility of having several correct answers, we ensured that only connectives with noncompeting functions were provided as answer choices. Considering that the task design controls for competing alternative functions of polyfunctional connectives and that only their dominant functions are targeted in the current study, we expect that the use of polyfunctional

connectives in the sentence cloze task should not be more difficult than the use of monofunctional connectives.

Moreover, as only two connectives per coherence relation type were tested by Tskhovrebova et al. (2022a), their results might have been influenced by the specific characteristics of the selected French connectives chosen to represent those relations. By conducting a study in a different language—i.e., in Russian—we ensured that the observed effects can be genuinely attributed to the tested factors rather than to the particularities of the selected connectives. In fact, previous research on connectives in adults usually demonstrates similar patterns across different languages, such as Dutch (Kamalski et al., 2008) and English (see, e.g., McNamara et al., 1996), and Dutch (Canestrelli et al., 2013) and Chinese (see, e.g., Wei et al., 2021).

Finally, it has been found that the mastery of connectives is variable among speakers of all ages (see, e.g., Volodina & Weinert, 2020, for the primary school years, Tskhovrebova et al., 2022a, for the teenage years,; Zufferey & Gygax, 2020a, for the adult years). To assess whether the use of connectives by Russian-speaking teenagers is modulated by their individual differences in a similar manner to French-speaking teenagers (Tskhovrebova et al., 2022a), two background measures of individual differences were examined, namely the degree of exposure to print and general vocabulary knowledge.

Connective-level predictors of the mastery of connectives by teenagers

By age 12, speakers can use and comprehend all types of common coherence relations such as addition, causality, concession, and contrast (e.g., Cain et al., 2005; McClure & Geva, 1983; Nippold et al., 1992). As measured in a cloze task, an appropriate use of the additive connective *and* (Cain et al., 2005), contrastive connective *but* (Cain et al., 2005; McClure & Geva, 1983) and concessive connective *although* (McClure & Geva, 1983) is achieved by age 10. The findings of Nippold et al. (1992) from cloze and sentence continuation tasks further demonstrated that, by age 12, teenagers had a high level of mastery of connectives encoding relations of addition (e.g., *furthermore*), consequence (e.g., *therefore*), concession (e.g., *however*), and contrast (e.g., *rather*).

Yet, not all connectives expressing a particular coherence relation are always used correctly. Some connectives may appear more difficult for young speakers because of the complexity of the coherence relation they signal (Sanders et al., 1992). However, this is not the only factor that may influence different levels of mastery of connectives. For instance, Crosson and Lesaux (2013) studied connectives that had different degrees of familiarity and encoded four coherence relations, namely, additive, adversative, causal, and temporal. Degrees of familiarity depended on the percentage of children who knew them and varied on the scale from 1 (most familiar) to 4 (least familiar). To give an example, among four adversative connectives (*but, or else, in contrast*, and *whereas*), *but* obtained the highest degree of familiarity and *whereas* the lowest. The results of the study revealed that, in a cloze sentence task, teenagers were more competent with more familiar connectives, regardless of the coherence relation signaled by these connectives.

It is also possible that the effect of connective familiarity may be interwoven with the effect of linguistic modality (oral versus written). In fact, exposure to oral speech happens as early as a child is born (or even prior to birth), while exposure to written language starts a lot later. It is mostly in secondary school that teenagers start to be extensively exposed to writing and become autonomous readers of a wide range of text genres (M. A. Nippold, 2004; Nippold, 2008). Moreover, there are important differences in the number and variety of connectives used in oral speech compared to writing. Not only is the repertoire of connectives used in the oral mode less diverse, but they are also used with less precise functions (Biber, 2006; Crible & Cuenca, 2017). For instance, Crible and Cuenca (2017) report that the connective *and* is used almost exclusively to mark additive relations in writing (90.76% of uses). In contrast, in speech, this connective is used with a greater number of functions such as causality, temporality and contrast, among which the signaling of the additive relation represents only 57.11% of occurrences. As a result, young speakers, who have not been exposed

long enough to the written modality are likely to perform less well on connectives that are usually used in writing than on those typically used in speech.

Some studies operationalized degree of familiarity with connectives, by measuring their frequency in corpora. These studies tried to dissociate the factors of modality and frequency (Nippold et al., 1992; Tskhovrebova, Zufferey, & Gygax, 2022; Zufferey & Gygax, 2020b). They did so by restricting their analysis to the connectives mostly used in writing, which allowed them to assess the effect of frequency independently from that of mode. The results demonstrated that frequency was still an important predictor of connective use even when the mode was controlled for. Yet, there were several gaps in these studies that refer to the role of frequency. The role of frequency in corpora for the competence with connectives was not tested in the work of Nippold et al. (1992), even though this study used frequency as a posthoc explanation of the results. Afterward, Tskhovrebova, Zufferey, and Gygax, (2022), and Zufferey and Gygax (2020b) corroborated this explanation by explicitly testing the factor of frequency. Both papers studied how four French connectives (*en outre* "in addition," *aussi* "therefore," *en effet* "because," and *toutefois* "however"), varying in frequency and bound to written language, were used in a cloze task by different groups of speakers. Zufferey and Gygax, (2020b) examined only high-school students aged 16 to 18, while Tskhovrebova, Zufferey, and Gygax, (2022) also assessed secondary-school students aged 12 to 15. Both studies showed that across all age groups the infrequent connectives *en outre* and *aussi* were more challenging in production tasks than the more frequent connectives *en effet* and *toutefois*.

These results may, however, have been biased by the polyfunctional nature of the connective *aussi*. In fact, this connective can signal both additive and consequence relations, depending on its position in the sentence. As a result, some participants erroneously chose *aussi* in the sentences where they were supposed to use the additive connective *en outre*, even though it is not possible to use *aussi* as an additive connective in the sentence-initial position (Roze et al., 2012). This error may have stemmed from the fact that *aussi* is more frequently used in the additive function than in the consequence one. Participants thus may have been guided by the probabilistic approach (Asr & Demberg, 2020) in spite of the syntactic constraints inherent to the connective *aussi*.

The fact that the polyfunctional connective *aussi* was problematic for teenagers may also suggest that polyfunctionality is a challenge with which older speakers have yet to learn to deal, when using connectives. The study of Zufferey et al. (2015) revealed, for example, that even adult native English speakers have difficulties discriminating between appropriate and inappropriate uses of the polyfunctional connective *while* that can express both temporal and contrastive relations. Speakers did not have trouble with evaluating the connective's dominant function (i.e., temporal meaning), but they were much less accurate when dealing with its secondary function (i.e., contrastive meaning). Hence, to overcome the design bias related to the polyfunctionality of some connectives, the present study assesses the use of both mono- and polyfunctional connectives, while ensuring that among answer options there are only connectives with noncompeting functions, to avoid the possibility of having several correct answers.

Student-level predictors of the mastery of connectives by teenagers

Intrinsic characteristics of connectives, such as coherence relation type, frequency, mode, and polyfunctionality, are not the only factors that potentially affect teenagers' competence with connectives. Individual differences in linguistic competence among teenagers may also have an impact on the skill of using appropriate connectives. There is evidence, for instance, that students following a pre-university curriculum are better at using connectives compared to pre-vocational students (Tskhovrebova, Zufferey, and Gygax, 2022; Zufferey & Gygax, 2020b). Moreover, pre-vocational students seem to depend more on connectives in a text, as their presence makes texts significantly more comprehensible for this group of speakers (van Silfhout et al., 2014). Finally, it has been demonstrated that age is another factor contributing to a better mastery of connectives, because

older teenagers tend to perform better in cloze and continuation tasks (Nippold et al., 1992) than younger children.

Research findings, showing that academic background (e.g., Tskhovrebova, Zufferey, & Gygax, 2022) and age (Nippold et al., 1992) play an important role in the use and comprehension of connectives, suggest that there may also be other sources accounting for the variation in the competence with connectives. The factor of academic background is rather broad and may combine a variety of skills that are more directly related to the mastery of connectives. It is possible that pre-university students who follow a more advanced school curriculum are more exposed to written texts and, thus, have richer overall vocabulary. As a result, they may also have a better knowledge of connectives, as these lexical items represent a separate area of the lexicon (Crosson & Lesaux, 2013) and appear in greater variety and with more precise functions in written texts (e.g., Crible & Cuenca, 2017). As a matter of fact, vocabulary size has been found to predict the use of connectives in a cloze test performed by adult native and nonnative French speakers (Wetzel et al., 2020) as well as by native French-speaking teenagers (Tskhovrebova et al., 2022a). Yet, only one study has examined the relation between vocabulary knowledge and the mastery of connectives in teenagers, and it remains an open question whether this pattern can be generalized to other languages. The study reported in this paper will determine whether and how the use of connectives is predicted by vocabulary size in Russian-speaking teenagers and in this way contribute to the research examining the development of the mastery of connectives between childhood and adult years.

Another predictor, which may be related to previously found variation in academic background is the degree of exposure to print, as pre-university students are supposed to have greater requirements in terms of readings than pre-vocational students, whose focus is on practical skills. Indeed, Zufferey and Gygax (2020a) demonstrated that the degree of exposure to print, as measured by the Author Recognition Test (ART; Stanovich & West, 1989), predicts the ability of adult speakers to identify correct and incorrect uses of connectives. In addition, Scholman et al. (2020) have established that the ART test is also an important factor predicting adults' ability to infer coherence relations indicating list relations from a different type of coherence signal, namely expressions of quantity. These findings thus point to a correlation between the competence to infer appropriate coherence relations and to use connectives in adult speakers. In addition, Tskhovrebova et al. (2022a) found that the ART was the second most important predictor of the performance in the connective-cloze test by French-speaking teenagers, following vocabulary size. However, there is still no evidence on the predictive power of this measure for the teenage cohort of Russian speakers. Replications (in other languages and learning contexts) are crucial to verify the role of exposure to print in the acquisition of connectives.

Finally, previous studies provide contrasting results on the role of age for the development of the competence with connectives. For primary school children, advancements in age have consistently been found to predict better comprehension and usage of connectives (e.g., Blything et al., 2015; Cain & Nash, 2011; Pyykkönen & Järvikivi, 2012). In contrast, the existing evidence on teenage years suggests that academic background is a stronger predictor of connective use in a sentence cloze task (Tskhovrebova, Zufferey, & Gygax, 2022; Zufferey & Gygax, 2020b). Therefore, there is a need to assess the role of age for the mastery of connectives also by Russian-speaking teenagers, especially in comparison to other measures of linguistic competence, namely vocabulary level and exposure to print.

The present study

The aim of this work was to continue the line of research assessing the factors that predict difficulties with the usage of discourse connectives. More specifically, this research examined a set of factors related to the properties of connectives themselves and to individual differences between speakers. Among the connective-level variables, we included type of coherence relation (addition, cause, concession, consequence, contrast, and temporality) and mode (oral versus written); and among student-level variables, we analyzed vocabulary size, level of exposure to print, and chronological age.

Our first research question was whether the connective mode (written or oral) predicts teenagers' performance in the cloze task, independently from the type of coherence relation. Our first hypothesis is that teenagers will probably have more difficulties in using written connectives than oral ones, independently of the coherence relation type, as massive exposure to written language comes later than exposure to oral language and all the main coherence relations are acquired by the end of the primary school years (age 11–12) (see, e.g., McClure & Geva, 1983).

Our second research question was whether vocabulary breadth and exposure to print predict connective use in a cloze test. In our third hypothesis, we predict that teenagers who have a larger vocabulary and those who are more exposed to print are more likely to use discourse connectives accurately, as connectives constitute a specific part of the lexicon (Crosson & Lesaux, 2013) and it is mostly through exposure to the written language that the widest variety of connectives can be acquired (see, e.g., Crible & Cuenca, 2017).

Our third research question was whether age predicts performance with connectives in a cloze task and whether its contribution is comparable to that of measures of linguistic experience. Our hypothesis was that, similar to previous studies on the teenage years (Tskhovrebova, Zufferey, & Gygax, 2022; Tskhovrebova et al., 2022a), the age factor will play a less important role for the mastery of connectives in comparison to individual differences in linguistic experience, as reflected by vocabulary level and exposure to print.

Method

Participants

One hundred and twenty-three native speakers of Russian, aged 11 to 17 (M_{age} = 13.68, SD = 1.87), participated in this study. Their native-level competence in Russian as well as absence of language disorders were validated by their teachers of Russian. The experiment was carried out in seven schools in Saint-Petersburg, in Russia, and included classes from the 5th to the 11th grade. A group of adults (N = 51, M_{age} = 33.37, SD = 8.08, Range 19–52) was also recruited for the experiment via the crowdsourcing platform Prolific© (Prolific, Oxford, UK, www.prolific.co) to establish the baseline of performance. All teachers, responsible for the classes of teenagers, as well as adult participants gave their informed consent for taking part in the experiment.

Materials

All the materials were created following the procedure described by Tskhovrebova et al. (2022a) in an experiment with French-speaking teenagers.

Sentence cloze test

Choice of connectives. Six types of common coherence relations (Sanders et al., 1992), namely addition, cause, concession, consequence, contrast, and temporality, were selected for this experiment. Each coherence relation was represented by two connectives – one that is more common in oral speech and one that is more prevalent in written language. Moreover, six connectives that were included in the task were polyfunctional, namely *hotia* "even if," *no* "but," *odnako* "however," *da i* "moreover," *vpročem* "nevertheless," *sledovatel'no* "therefore," and another six were monofunctional, namely *potomu čto* "because," *tak čto* "so," *kak tolko* "as soon as," *krome Togo* "moreover," *tak kak* "because," *edva* "as soon as." The number of functions that each connective can encode was determined based on the dictionaries of Yefremova (2000) and Yevgen'eva (1999). Moreover, we also conducted a corpus analysis to trace the dominance of the tested functions for these polyfunctional connectives. For each connective, we annotated 50 sentences, randomly extracted from the oral subcorpus of the *Russian National Corpus* (https://ruscorpora.ru; Grishina & Savchuk, 2009), as well as 50 sentences extracted from the written subcorpus. The results of the corpus analysis showed that for the majority of the

Table 1. Number of occurrences of different functions of the polyfunctional connectives in written and oral corpora.

	Concession	Consequence	Contrast	Addition	Total
da i "moreover"			9	91	100
Oral			2	48	50
Written			7	43	50
hotia "even if"	85		15		100
Oral	44		6		50
Written	41		9		50
no "but"	6		80	14	100
Oral	4		39	7	50
Written	2		41	7	50
odnako "however"	8		90	2	100
Oral	5		43	2	50
Written	3		47		50
sledovatel'no "therefore"		79		21	100
Oral		40		10	50
Written		39		11	50
vpročem "nevertheless"	35		40	25	100
Oral	8		32	10	50
Written	27		8	15	50

polyfunctional connectives, the tested function was dominant both in written and oral corpora (see Table 1 for the distribution of different functions). It appears that for the connective *vpročem* "nevertheless," it is not possible to distinguish one dominant function, as the concessive use tends to be more frequent in the written mode, and the contrastive function is more frequent in the oral mode.

To determine which connectives were more bound to oral speech and which ones to written language, we conducted a corpus analysis of connective frequencies and administered a questionnaire to gauge native speakers' judgments. We calculated the connective frequencies in oral speech based on the *Russian National Corpus*, as it is large (13.4 million words) and contains speech from a wide variety of genres and degrees of formality, such as everyday conversation and public speech. The connective frequencies in writing were calculated based on the written subcorpora of the *Russian National Corpus*, including journalistic, literary, scientific, and technical texts. Those connectives that had a higher frequency in oral than in written corpora were classified as *oral*; and those with a higher frequency in the written subcorpus were categorized as *written*.

In addition to the corpus study, we recruited 109 adult native Russian speakers online to verify whether each of the selected connectives was common for an informal oral conversation, such as the one at a dinner with friends. The participants had to make their evaluation on a scale from 0 to 20. If they thought that a connective was never used in informal oral conversation, they were asked to choose the answer 0; and if they believed that it was used in such contexts very often, they were asked to choose the answer 20. For each pair of connectives representing the same coherence relation, the connective with a higher total was labeled as *oral* and that with a lower one as *written*. The results from the judgment test were congruent with the categorization based on the corpus analysis. Participants who fulfilled the judgment task did not take part in the main experiment. Table 2 shows the distribution of the 12 selected connectives per coherence relation, modality, and polyfunctionality.

Design of the cloze task. Participants were asked to fill in a blank between two sentences with a correct connective, making a choice out of four options. The blank was delimited with double slashes "//_____//" instead of punctuation marks so that punctuation on the border between two sentences did not affect the choice of a connective. There were 60 pairs of sentences in the task. Each coherence relation was represented by 10 items, half of which tested a connective typical to oral speech and another half the one mostly used in writing. We tested oral and written modalities separately to avoid more common oral connectives being always selected instead of written ones. To do so, we presented only oral connectives as

Table 2. Distribution of connectives per type of coherence relation and modality 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	Poly-functionality
Addition	oral	*da i*[a]	moreover (less formal)	14.09 (5.24)	232.18	163.89	+
	written	*krome Togo*	moreover (more formal)	10.48 (5.09)	93.16	307.10	–
Cause	oral	*potomu čto*	because (less formal)	17.67 (3.16)	2565.03	453.94	–
	written	*tak kak*	because (more formal)	12.06 (5.28)	191.90	286.23	–
Concession	oral	*hotia*	even if	16.02 (4.09)	630.96	536.00	+
	written	*vpročem*	nevertheless	8.77 (5.35)	19.00	204.45	+
Consequence	oral	*tak čto*	so	15.96 (3.93)	474.69	192.43	–
	written	*sledovatel'no*	therefore	7.63 (5.19)	28.51	60.92	+
Contrast	oral	*no*	but	18.68 (2.39)	5999.09	4427.84	+
	written	*odnako*	however	9.16 (5.51)	51.02	738.67	+
Temporality	oral	*kak tolko*	as soon as (less formal)	12.81 (4.62)	57.53	54.99	–
	written	*edva*	as soon as (more formal)	6.92 (4.78)	10.74	107.71	–

[a]Russian connectives were transliterated from the Cyrillic to the Latin alphabet for convenience throughout the article.

answer options in sentences targeting oral connectives and only written connectives in sentences targeting written connectives. Examples (3) and (4) illustrate how this principle was applied to the relation of consequence.

(3) The correct answer: written connective *sledovatel'no* 'therefore'
Саша пропустил много лекций //_____// ему будет непросто на экзамене.
'Sasha has missed a lot of lectures //_____// he will have a hard time at the exam.
'Answer options: (a) *vpročem* 'nevertheless'; (b) *krome Togo 'moreover (more formal)'*; (c) *sledovatel'no* 'therefore'; (d) *edva 'as soon as (more formal)'*

(4) The correct answer: oral connective *tak čto* 'so'
Маша не спала всю ночь //_____// на утро у нее сильно болела голова.
'Masha stayed up all night //_____// she had a bad headache in the morning.
'Answer options: (a) *hotia 'even if'*; (b) *da i 'moreover (less formal)'*; (c) *tak čto* 'so'; (d) *kak tolko 'as soon as (less formal)'*

To verify whether the linguistic context in which connectives were used was well suited for all the tested connectives, we asked a different group of 40 adult native Russian speakers, recruited via an online platform (Prolific, Oxford, UK, www.prolific.co), to judge the acceptability of the task sentences in the Russian language on a scale from 0 to 10, where 0 stood for *absolutely not acceptable* and 10 stood for *absolutely acceptable*. To the 60 items from the cloze test with correctly inserted connectives, we added 24 fillers. Half of the fillers included wrong connectives, as in (5), and another half included lexico-grammatical mistakes, such as wrong usage of phrasal expressions (6) and verb government errors (7). In the example (6), for instance, the phrasal expression *язык не поворачивался* "the tongue did not turn" was replaced by the wrong expression *язык не поднимался* "the tongue did not rise." The sentence (7), in turn, included an error in the government of the verbal expression *уделять внимание* "pay attention" that should be followed by the noun *child* in the dative case (*ребёнку*). Instead, it was followed by the noun *child* in the accusative case with the preposition *на* "on" (*на ребёнка*).

(5) Пошёл снег, потому что мы зашли домой.
"It started snowing because we came home."

(6) Маша была так напугана, что у неё язык не поднимался ничего сказать в ответ.
"Masha was so frightened that she had no tongue to say anything back."

(7) Антонина не достаточно уделяет внимание на ребёнка из-за большой занятости на работе.
"Antonina doesn't pay enough attention to the child because she is very busy at work."

The results of the acceptability judgment task showed that the overall acceptability rate of the cloze test items (M = 8.33, SD = 2.36, Range 0–10), as well as the acceptability of the specific items testing oral (M = 8.72, SD = 2.02, Range 0–10) and written (M = 7.95, SD = 2.61, Range 0–10) connectives were quite high. This suggests that both types of connectives were tested in a context that was suitable for them.

We also ensured that there was only one possible answer for each item, as connectives whose primary or secondary function could interfere with the targeted connective were not included in the choice of answers. For instance, in sentences testing concessive connectives *vpročem* "nevertheless" and *hotia* "even if," we did not propose as answer options connectives *odnako* "however" and *no* "but" that can signal contrastive and concessive relations. We calculated the score for this task as a proportion of correct answers per connective.

Vocabulary test

The participants' vocabulary size was measured with a newly developed vocabulary level test based on Nation and Beglar (2007). The task involved choosing out of six options the word that corresponded best to a given definition. There were four groups of words (30 items each), selected from frequency lists of the *Russian National Corpus* (https://ruscorpora.ru; Lyashevskaya & Sharoff, 2009) and representing the first, second, third, and fourth 5,000-word families. The participants fulfilled the task by starting from the first group of words, having the highest frequency, and by finishing with the fourth group of words, having the lowest frequency. Among the words included in each group, there were 18 nouns, 6 verbs, and 6 adjectives. The reliability of the test, as measured by Cronbach's alpha, was high for both teenagers, .99 [.986–.991],[1] and adults, .91 [.85–.93]. Hence, the total vocabulary score was computed as the percentage of correct responses per participant.

Author recognition tests

To assess the teenagers' level of exposure to print, we administered an author recognition test (ART) developed specifically for this study. The new version of the ART (*ART-RU-CL*) included 40 names of classic authors and 40 filler nonauthor names, that were presented in a random order. The classical authors' names were chosen based on the classifications provided by four national bookstore chains. Participants were asked to choose all the authors' names that they knew. To avoid guessing, the participants were instructed to check only those names about which they were sure, as not all the names belonged to real authors, and they would lose one point per each wrongly selected name. One point was attributed for each correctly checked author, and −1 for each wrongly chosen one. To calculate the final score, we made a sum of correct and wrong answers, where the minimum total score was *−40* and the maximum was *40*.

Another version of the ART (*ART-RU*) was developed for the group of adults as this measure is sensitive to the age of participants (see, e.g., Allen et al., 1992; Cunningham & Stanovich, 1990). Our test followed design principles developed by Stanovich and West (1989) in the original ART and included the names of modern and contemporary authors who won literary prizes or are bestselling. The test structure and the score calculation procedure were the same as in the ART-RU-CL. The reliability of both ART tests was high, as indicated by their Cronbach's alphas greater than .90 (ART-RU-CL: .94 [.92–.95]; ART-RU: .92 [.86–.94]).

Finally, all the participants also provided a subjective evaluation of their exposure to print. More specifically, they had to evaluate their reading habits on the scale from 0 to 10, where 0 stood for *never* and 10 for *every day*.

Procedure

All the participants performed the tasks in the same order, starting with the connective cloze task and then proceeding to the author recognition test and the vocabulary level test. They could not return to previous questions and make changes as soon as they clicked the button leading to the next question. Teenagers fulfilled all the tasks online via a link that they received directly on their classroom computers. It took them approximately one hour to finish the whole test battery. Adults also completed the tasks online, but via the Prolific website (https://www.prolific.co), and spent around 40 minutes on it.

Analysis

We used a generalized mixed-effects logistic regression model to analyze binary responses (right or wrong) on the connective test in the R software (R Core Team, 2012). To examine whether there was a general difference between teenagers and adults in the performance on the connective task, we first analyzed the results of all participants together. After a global analysis, we separately analyzed the results of teenagers and adults to assess the role of the predictors of individual variation within each group. The measures of inter-individual variation were not included in the analysis for all participants because different versions of the ART were used to assess the degree of exposure to print in teenagers and adults.

First, we centered all the predictors of individual variation and then created a full model with the *glmer* function of the *lme4* package (Bates et al., 2015), including all the relevant variables. The outputs of the full models for convenience are provided in the Appendix on the OSF repository. The model for all participants included group (adults versus teenagers) and connective mode (oral versus written). Only two predictors were added in this analysis, as we wanted to make a general comparison between the level of performance with connectives of the two age groups.

The model for teenagers included degree of exposure to print (ART-RU-CL), vocabulary level, subjective evaluation of exposure to print, connective mode, coherence relation, polyfunctionality (monofunctional versus polyfunctional), and age. When we checked our data for multicollinearity, we observed that vocabulary score correlated with ART-RU-CL (*rho* = .73 [.64, .81], $p < .001$) and with age (*rho* = .33 [.16, .49], $p < .001$). Moreover, age and ART-RU-CL also correlated with each other (*rho* = .33 [.16, .48], $p < .001$). To avoid multicollinearity, we first residualized vocabulary scores by ART-RU-CL and age, and then age by ART-RU-CL, using the *umx_residualize* function of the *umx* package (Bates, 2021).

In the full model for adults, there were the same variables as in the one for teenagers, except for the ART-RU that was included instead of the ART-RU-CL, adapted for teenagers. Similar to the teenagers' data, age correlated with ART-RU-CL (*rho* = .30 [.01, .53], $p < .001$); and vocabulary score correlated with ART-RU (*rho* = .55 [.31, .72], $p < .001$) and age (*rho* = .43 [.17, .64], $p < .001$). To avoid multicollinearity, we also residualized vocabulary scores by ART-RU and age, as well as age by ART-RU.

After having built the full models, with the *drop1* function of the *stats* package (R Core Team, 2012), we automatically selected only those predictors that were relevant for our reduced models. When the factors of mode and group or mode and coherence relation were selected among other relevant predictors for the reduced model, we checked whether adding an interaction between these factors improved the final model's fit. Comparison between the models without and with an interaction was done with the *anova* function of the *stats* package. The final reduced models were returned with the *summary* function of the *lmerTest* package (Kuznetsova et al., 2017). The significance threshold was set at $p < .05$. In addition, when relevant, we performed a post hoc pairwise comparison between the tested predictors, using the *lsmeans()* function of the *emmeans* package in R (Lenth, 2020).

Treatment contrasts were applied to all unordered factors across all three analyses. Oral mode was set as reference to compare the two modalities, because speakers start to be exposed to oral

speech much earlier than to written language and should master it better. For the factor of group, adults were chosen as the baseline, as this group was assumed to include speakers with the highest level of competence. Age was treated as a categorical variable (adults versus teenagers) in the analysis for all participants, as we wanted to examine whether there was a general difference between teenagers and adults in the performance on the connective task. In the separate analyses for teenagers and adults, age was treated as a continuous variable and was centered just like other continuous variables. The causal relation was set as reference level for comparing the scores associated to different coherence relations, as speakers have a cognitive bias toward causality in the absence of other explicit cues (Sanders, 2005) and should not experience difficulties with this relation (see, e.g., Cain & Nash, 2011).

In the separate models for teenagers and adults, we also performed a random forest analysis based on the predictors included in the final reduced model, to compare the impact of each relevant predictor on the dependent variable (i.e., correctness of responses in the cloze task; Strobl et al., 2009). We chose to complement the regression analysis with this method because it can deal with highly correlated variables (such as our measures of individual variation), as it does not rely on any assumptions about the distribution of data. This method is also robust because it computes the importance of each variable on the basis of a large number of regression trees (Strobl et al., 2009).

Results

All participants

The final model for all participants included fixed effects of group and mode, as well as a two-way interaction between group and mode, as adding an interaction between group and mode improved the models' fit ($x^2(1) = 4.63$, $p = .031$). The output of the statistical analysis (see Table 3) showed that teenagers scored significantly lower than adults on oral connectives. Moreover, adults on average

Table 3. Output of the final model for all participants.

Variable	Estimate	*SE*	*Z*	Pr(>|z|)
All participants				
(Intercept)	5.43	0.40	13.73	<0.001
Teenagers	−2.64	0.42	−6.32	<0.001
Written mode	−1.38	0.31	−4.46	<0.001
Teenagers*Written mode	0.58	0.28	2.08	0.038

Note. Because the output of the final model includes the same predictors as the initial model, we do not report the initial model in the Appendix.

Table 4. Mean accuracy score per connective among teenagers and adults.

				Teenagers		Adults	
Relation	Mode	Connective	Translation in English	*M*	95% CI	*M*	95% CI
Addition	oral	*da i*	moreover	.81	[.71, .91]	.96	[.92, 1.01]
	written	*krome Togo*	moreover	.82	[.72, .91]	.96	[.91, 1.01]
Cause	oral	*potomu čto*	because	.87	[.71, .91]	.98	[.95, 1.02]
	written	*tak kak*	because	.82	[.73, .92]	.99	[.97, 1.01]
Concession	oral	*hotia*	even if	.83	[.73, .92]	.99	[.96, 1.02]
	written	*vpročem*	nevertheless	.71	[.60, .83]	.92	[.86, .99]
Consequence	oral	*tak čto*	so	.81	[.71, .91]	.99	[.97, 1.01]
	written	*sledovatel'no*	therefore	.75	[.64, .86]	.92	[.85, .99]
Contrast	oral	*no*	but	.85	[.76, .94]	.99	[.98, 1.01]
	written	*odnako*	however	.80	[.70, .90]	.98	[.94, 1.01]
Temporality	oral	*kak tolko*	as soon as	.88	[.80, .96]	1	[1, 1]
	written	*edva*	as soon as	.66	[.54, .78]	.93	[.87, 1]

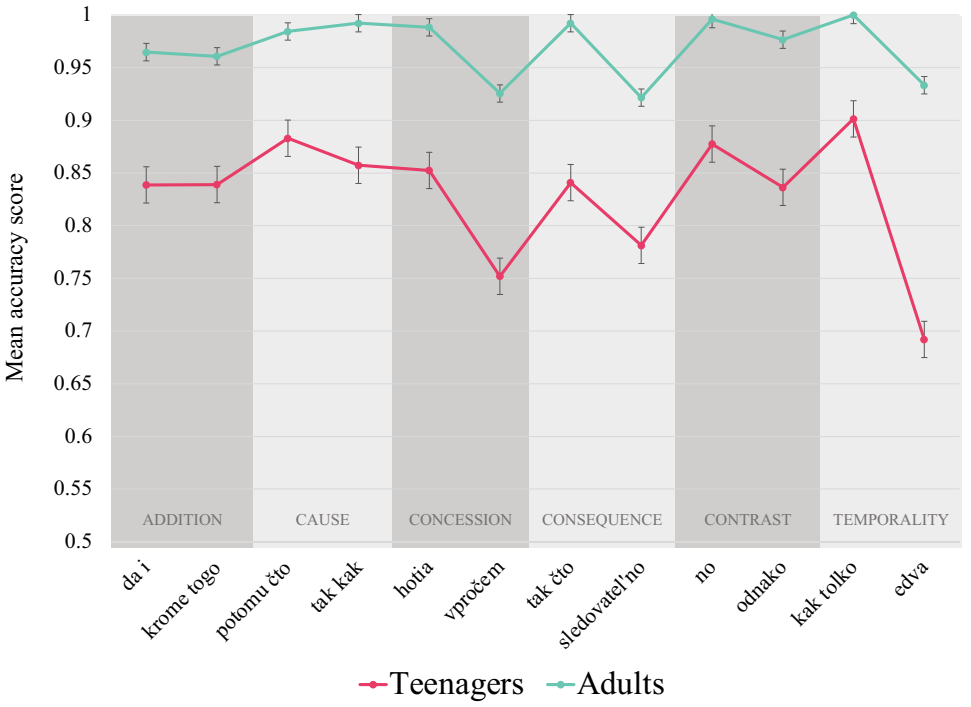

Figure 1. Distribution of mean scores per connective in sentence cloze task among teenagers and adults. *Note.* The translations (and modes) of the Russian connectives are as follows: *da i* "moreover (oral)," *krome Togo* "moreover (written)," *potomu čto* "because (oral)," *tak kak* "because (written)," *hotia* "even if" (oral), *vpročem* "nevertheless"(written), *tak čto* "so" (oral), *sledovatel'no* "therefore" (written), *no* "but" (oral), *odnako* "however" (written), *kak tolko* "as soon as (oral)," *edva* "as soon as (written)."

scored lower for written connectives than for the oral ones. A significant interaction between group and mode seems to suggest that the difference in the scores between oral and written connectives was smaller in the group of teenagers. However, teenagers overall scored quite high, namely between .80 and .88 for the majority of connectives, except for the written connectives *vpročem* "nevertheless" (M = .71, [.60, .83]), *sledovatel'no* "therefore" (M = .75, [.64, .86]), and *edva* (M = .66, [.54, .78]). The general accuracy level of adults, in contrast, was above 90% and varied between .92 and 1. See Table 4 for the mean scores per connective within each age group and Figure 1 for a visual representation of the results.

Teenagers

An automatic step-back analysis of data showed that the performance in the connectives test within the group of teenagers was mostly predicted by vocabulary level, exposure to print, age, coherence relation, and mode. Including an interaction between mode and coherence relation improved the fit of the final model ($x^2(5) = 27.63$, $p = < .001$) (see Table 5). Polyfunctionality and subjective exposure to print were not revealed to be significant predictors of teenagers' performance in the connective task.

A higher vocabulary level accounted for an estimated increase of 6.99 ± 0.61 *SE* in the connectives task, higher scores in ART-RU-CL were associated with an estimated increase of 1.96 ± 0.14 *SE*, and age accounted for an increase of 2.29 ± 0.77 *SE* estimates (see predictors 1, 2, and 3 in Table 5). Moreover, the oral additive connective *da I* "moreover" (M = .81, [.71, .91]; predictor 5) and the oral consequence connective *tak čto* "so" (M = .81, [.71, .91]; predictor 7) had significantly lower scores than the oral causal connective *potomu čto* (M = .87, [.71, .91]). Teenagers also scored significantly

Table 5. Output of the final reduced model for teenagers.

Variable	Estimate	*SE*	*Z*	Pr(>|z|)
Teenagers				
(Intercept)	−2.56	0.44	−5.83	<0.001
1. Vocabulary test**	6.99	0.61	11.46	<0.001
2. ART-RU-CL*	1.96	0.14	14.05	<0.001
3. Age*	2.29	0.77	2.98	0.003
MODE				
4. Written	−0.49	0.30	−1.64	0.102
COHERENCE RELATION				
5. Addition	−0.63	0.30	−2.12	0.034
6. Concession	−0.44	0.30	−1.47	0.142
7. Consequence	−0.66	0.30	−2.23	0.026
8. Contrast	−0.16	0.30	−0.53	0.593
9. Temporality	0.20	0.30	0.64	0.521
MODE*COHERENCE RELATION				
10. Written*Addition	0.56	0.42	1.34	0.181
11. Written*Concession	−0.63	0.41	−1.53	0.126
12. Written*Consequence	−0.07	0.41	−0.16	0.874
13. Written*Contrast	−0.09	0.42	−0.22	0.828
14. Written*Temporality	−1.68	0.42	−4.03	<0.001

*Centered values.
**Centered and residualized values.

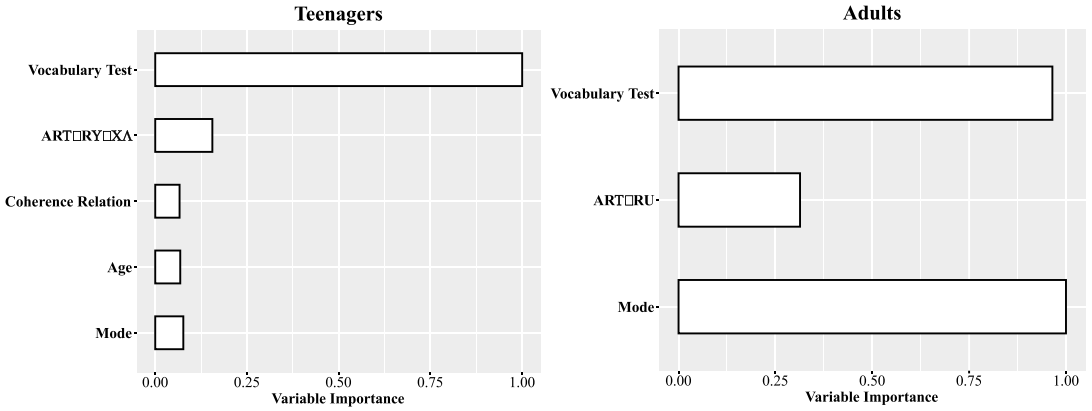

Figure 2. The impact of each predictor variable on the dependent variable according to the random forest analyses for teenagers and adults.

lower for the written temporal connective *edva* "as soon as" (*M* = .66, [.54, .78]; predictor 14) than for the oral causal connective *potomu čto* (see Table 5). Furthermore, pairwise comparison also revealed that *edva* "as soon as" received the lowest score compared to almost all the other written connectives, such as *krome Togo* "moreover" (log odds ratio = −1.42, *SE* = 0.28, *p* = < .0001), *kak tak* "because" (log odds ratio = −1.49, *SE* = 0.29, *p* = < .0001), *odnako* "however" (log odds ratio = −1.24, *SE* = 0.28, *p* = < .0001), and *sledovatel'no* "therefore" (log odds ratio = −0.76, *SE* = 0.28, *p* = .0063).

Finally, a significant interaction between mode and coherence relation revealed that the effect of mode was more pronounced for some coherence relations than for the others. For instance, teenagers produced significantly lower scores for the written mode only for the relations of concession (log odds ratio = −1.12, *SE* = 0.29, *p* = < .0001) and temporality (log odds ratio = −2.17, *SE* = 0.29, *p* = < .0001).

The random forest analysis (prediction accuracy of 90%) demonstrated that the most important predictor of performance with connectives by teenagers was vocabulary level, outranking exposure to print, age, coherence relation and mode (Figure 2). The scores on the vocabulary test and the ART across teenagers and adults are reported in Table 6.

Table 6. Descriptive statistics for background measures among teenagers and adults.

Teenagers		Adults		
M (SD)	Observed range	*M* (*SD*)	Observed range	Possible range
Vocabulary size				
.73 (.29)	.12–1	.96 (.06)	.79–1	0–1
Author recognition test*				
13 (10.01)	−6–40	10.71 (7.76)	0–32	−40–40
Subjective exposure to print				
5.95 (2.37)	1–10	6.63 (2.07)	2–10	0–10

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

Table 7. Output of the final reduced model for adults.

Variable	Estimate	*SE*	*z*	Pr(>|z|)
Adults				
(Intercept)	3.99	0.57	7.06	<0.001
Vocabulary test**	19.76	5.50	3.59	<0.001
ART-RU*	0.71	0.20	3.58	<0.001
MODE				
Written	−1.51	0.46	−3.29	0.001

*Centered values.
**Centered and residualized values.

Adults

The final reduced model for the group of adults included vocabulary level, exposure to print, and mode (see Table 7). Polyfunctionality, coherence relation type, age, and subjective exposure to print were not revealed to be significant predictors of the performance in the cloze test. The same reference levels as in previous analyses were applied to the unordered factor of mode. Higher scores on the vocabulary level test were associated with an estimated increase of 19.76 ± 5.50 *SE* in the sentence cloze task; and a greater degree of exposure to print, as measured by the ART-RU, accounted for an estimated increase of 0.71 ± 0.20 *SE*. In contrast, the written mode accounted for a decrease of 1.51 ± 0.46 *SE* estimates (see Table 7), meaning that written connectives were on average more challenging than the oral ones. The random forest analysis had a prediction accuracy of 97% and revealed that, for the group of adults, the most important predictors of accurate use of connectives in the cloze test were vocabulary level and mode, followed by exposure to print (see Figure 2).

Discussion

Factors related to the properties of connectives

This study set out to explore the factors influencing the acquisition of Russian discourse connectives in the teenage years. In line with the earlier study on French (Tskhovrebova et al., 2022a), our experiment demonstrated that, when used in a monofunctional context, connectives do not pose much difficulty for Russian-speaking teenagers, at least for the 12 connectives included in the present experiment. However, it is possible that the high performance in the cloze test was due to the task design, in which participants had to fill in blanks between pairs of sentences and not in a more ecological context of texts. Indeed, it was shown by Tskhovrebova, Zufferey, and Gygax, (2022) that it is cognitively more challenging for both teenagers and adults to apply an appropriate connective within a short text than between two sentences.

Moreover, only the dominant functions of the polyfunctional connectives were targeted by the design of the present study. In future studies, it will therefore be useful to increase the challenge of the task and to examine the use of nondominant functions of connectives as well as the use of infrequent

connectives, expressing rare coherence relations, in more realistic contexts. Although the current task modality was probably not particularly challenging for teenagers, their scores were still lower than those of adults. This result supports studies by Berman (2004) and Nippold (2008) and suggests that adult competence to use connectives is not completely acquired by age 18 and continues to develop far beyond the high school years.

Not all connectives, however, were mastered equally well by teenagers. There was a significant interaction between a coherence relation type and mode on the performance of teenagers in the cloze test. The written concessive and temporal connectives *vpročem* "nevertheless" and *edva* "as soon as" received lower scores than their oral counterparts. This result is in line with the prior finding on French-speaking teenagers (Tskhovrebova et al., 2022a) and supports our first hypothesis on the role of modality, indicating that lack of sufficient exposure to the written modality may be at the heart of the lower performance, at least with some written connectives. To improve competence with this type of connectives, teenagers should have more training at school with different registers of written language as well as with connectives that most often appear in this mode. However, because only one connective per mode and coherence relation type was assessed in this study, it is important to mention that the results apply only to these specific 12 connectives.

The score for the temporal written connective *edva* "as soon as" was particularly low in the group of teenagers. It is difficult to propose a solid explanation for this finding. Our tentative suggestion would be that this word can also be used as an adverb, translated into English as *barely* and *slightly*. Therefore, these nonconnective meanings may interfere and mislead young speakers in their attempts to match an appropriate connective with a temporal coherence relation. However, a more comprehensive examination of how nonconnective meanings may affect the performance with the connective functions should be conducted in future studies.

Student-level predictors of the mastery of connectives

The results have demonstrated that, similarly to the findings for French-speaking teenagers (Tskhovrebova et al., 2022a), vocabulary level was the most important predictor of the varying competence with connectives in the cloze task. This result suggests that vocabulary knowledge plays an important role in the use of connectives starting from early teenage years and during adulthood, even though connectives are not typical lexical items, as they encode procedural rather than (or along with) conceptual meaning (see, e.g., Wilson, 2011).

Another factor that predicted differences in performance on the cloze task was the level of exposure to print, as assessed by the newly created Russian versions of the ART. This finding complements the existing research in adults (Scholman et al., 2020; Zufferey & Gygax, 2020a) and French-speaking teenagers (Tskhovrebova et al., 2022a), showing that the ability of Russian-speaking teenagers and adults to use connectives also varies according to the degree of exposure to print, starting as early as at age 11. It should be noticed that even though the ART does not measure the actual number of books that a person has read in their life, it has been shown to be indicative of a level of general linguistic competence, including sentence-processing skill (Acheson et al., 2008), world and vocabulary knowledge (see, e.g., Stanovich et al., 1995), as well as metacognitive competence (McBride-Chang & Chang, 1995).

Because the ART is a proxy of exposure to print and reflects a complex set of language skills, this may be a reason why it explained variation in the performance with both written and oral connectives. Furthermore, it should be mentioned that the comparison of the effects from the ART on the use of connectives by teenagers and adults must be done with caution, as two different versions were used to measure the degree of exposure to print in the two age groups of participants. In general, the question of how to handle the scores from the ARTs, adapted to different age groups of speakers, should be explored in more detail in future research.

Although the ART test is an indirect measure of exposure to print, it was better suited for assessing differences in the use of discourse connectives than the subjective evaluation of exposure to print. The

reason for such a difference between indirect and direct measures of exposure to print is that self-report tests are more prone to socially desirable answers (see, e.g., Echols et al., 1996; Wimmer & Ferguson, 2022). Therefore, in future work, the ART tests should be preferred to subjective evaluations of exposure to print.

Finally, age was found to be another factor predicting better performance in a connective cloze task, but only for the group of teenagers. However, the random forest analysis demonstrated that age did not play an important role as lexicon size and degree of exposure to print. This finding is not in line with previous studies on younger children, where mastery of connectives was strongly predicted by age (Blything et al., 2015; Pyykkönen & Järvikivi, 2012), but it corroborates the results of a similar experiment on French-speaking teenagers (Tskhovrebova et al., 2022a). Although this interesting result needs further replications within and across languages, the finding that age plays a less important role than the measures of individual variation both for French- and Russian-speaking teenagers may suggest that later linguistic development is qualitatively different and follows a slower pace than early language acquisition (Nippold, 1993). Appropriate use of a wide variety of connectives, which happens at a later developmental stage, probably requires additional effort and extensive input, especially through reading, on top of age-related cognitive maturation. In addition, this result may also highlight the fact that chronological age is an indirect measure of various competences, such as cognitive skills and language experience, that are subject to individual variation (Kidd et al., 2018). Indeed, the factor of age was not relevant at all for the group of adult speakers, whose mastery of connectives probably does not further develop with biological age.

Conclusion

Our study emphasizes the importance of research on linguistic development in teenage years, a period that has received relatively little attention in language acquisition research. The present results corroborate and strengthen general findings from a similar study on French-speaking teenagers (Tskhovrebova et al., 2022a) and show that, by age 18, Russian-speaking teenagers have a high command of 12 connectives signaling most common coherence relations, belonging to the written and the oral modes, and having one or several functions. In our experiment, we also reported a strong individual variation among speakers in the ability to match connectives with an appropriate coherence relation.

The use of connectives in a cloze task was strongly predicted by teenagers' vocabulary size and level of exposure to print and more marginally by speakers' age, connective modality, and coherence relation type. This may suggest that, at least starting from 11, age and intrinsic properties of connectives matter less for their mastery than general linguistic experience, as measured by vocabulary level and exposure to print. Moreover, exposure to print and vocabulary size continue to explain individual differences in the performance with connectives even during the adult years. Therefore, school curricula should support and promote students' exposure to written texts, which will enable them to expand their vocabulary size and become more familiar with the use of a greater number of connectives, thereby enhancing reading comprehension skills and general academic performance.

Note

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

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Swiss National Science Foundation under Grant [100012_184882].

ORCID

Ekaterina Tskhovrebova http://orcid.org/0000-0002-1116-3206
Sandrine Zufferey http://orcid.org/0000-0002-5403-6709
Elena Tribushinina http://orcid.org/0000-0003-1688-7307

Data availability statement

All materials, data, and code of this study are accessible on the OSF repository (https://osf.io/5khsr/?view_only=4ac5b6a132044d29ba002127afff7dcf).

References

Acheson, D., Wells, J., & MacDonald, M. (2008). New and updated tests of print exposure and reading abilities in college students. *Behavior Research Methods*, *40*(1), 278–289. https://doi.org/10.3758/BRM.40.1.278

Allen, L., Cipielewski, J., & Stanovich, K. E. (1992). Multiple indicators of childrens 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

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*(4), 978–1021. https://doi.org/10.1111/lang.12365

Baştuğ, M. (2014). the structural relationship of reading attitude, reading comprehension and academic achievement. *International Journal of Social Sciences and Education*, *4*(4), 931–946.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bates, T. (2021). *umx: A helper package for structural equation modeling in OpenMx*. University of Edinburgh. https://doi.org/10.5281/zenodo.10937

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

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. John Benjamins. https://doi.org/10.1075/scl.23

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

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

Cain, K., Patson, N., & Andrews, L. (2005). Age- and ability-related differences in young readers' use of conjunctions. *Journal of Child Language*, *32*(4), 877–892. https://doi.org/10.1017/S0305000905007014

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

Crible, L., & Cuenca, M. (2017). Discourse markers in speech: Characteristics and challenges for annotation. *Dialogue and Discourse*, *8*(2), 149–166. https://doi.org/10.5087/dad.2017.207

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*(3), 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

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*(7/8), 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*(2), 296–304. https://doi.org/10.1037/0022-0663.88.2.296

Grishina, E., & Savchuk, S. (2009). Korpus ustnyh tekstov v NKRJa: Sostav i struktura [Corpus of oral texts in RNC: Content and structure]. In V. A. Plungjan, E. V. Rahilina, & T. I. Reznikova (Eds.), *Nacional'nyj korpus russkogo jazyka: 2006—2008. Novye rezul'taty i perspektivy* [National corpus of the Russian language: 2006–2008. New results and perspectives] (pp. 129–149). NestorIstorija.

Kamalski, J., Sanders, T., & Lentz, L. (2008). Coherence marking, prior knowledge, and comprehension of informative and persuasive texts: Sorting things out. *Discourse Processes*, *45*(4–5), 323–345. https://doi.org/10.1080/01638530802145486

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends Cogn Sci*, *22*(2), 154–169. https://doi.org/10.1016/j.tics.2017.11.006

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

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. https://doi.org/10.18637/jss.v082.i13

Lenth, R. (2020). *emmeans: Estimated marginal means, aka least-squares means. R package version 1.5.1*. Vienna: R Core Team. https://CRAN.R-project.org/package=emmeans

Lyashevskaya, O., & Sharoff, S. (2009). *Chastotny slovar sovremennogo russkogo yazyka* [Frequency dictionary of contemporary Russian]. Azbukovnik.

McBride-Chang, C., & Chang, L. (1995). Memory, print exposure, and metacognition: Components of reading in Chinese children. *International Journal of Psychology*, *30*(5), 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*(4), 411–432. https://doi.org/10.1080/01638538309544575

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14*(1), 1–43. https://doi.org/10.1207/s1532690xci1401_1

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

Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*(7), 9–13.

Nippold, M. (1993). Developmental markers in adolescent language: Syntax, semantics, and pragmatics. *Language, Speech, and Hearing Services in Schools*, *24*(1), 21–28. https://doi.org/10.1044/0161-1461.2401.21

Nippold, M. (2008). *Later language development: School-age children, adolescents, and young adults* (3rd ed., 2nd printing). PRO-ED.

Nippold, M. A. (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., 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*(1), 108–118. https://doi.org/10.1044/jshr.3501.108

Pyykkönen, P., & Järvikivi, J. (2012). Children and situation models of multiple events. *Developmental Psychology*, *48*(2), 521–529. https://doi.org/10.1037/a0025526

R Core Team. (2012). *R: A language and environment for statistical computing* [Computer software manual]. http://www.R-project.org/

Roze, C., Danlos, L., & Muller, P. (2012). LEXCONN: A French lexicon of discourse connectives. *Discours*, *10*, 1–15. https://doi.org/10.4000/discours.8645

Salihu, 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/salihu.pdf

Sanders, T. (2005). Coherence, causality and cognitive complexity in discourse. In *Proceedings/ Actes SEM-05, First International Symposium on the Exploration and Modelling of Meaning*.

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

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

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

ter Beek, 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

Tskhovrebova, E., Zufferey, S., & Gygax, P. (2022). Individual variations in the mastery of discourse connectives from teenage years to adulthood. *Language Learning*, *72*(2), 412–455. https://doi.org/10.1111/lang.12481

Tskhovrebova, E., Zufferey, S., & Tribushinina, E. (2022a). French-speaking teenagers' mastery of connectives: The role of vocabulary size and exposure to print. *Applied Psycholinguistics*, *43*(5), 1141–1163. https://doi.org/10.1017/S0142716422000303

Tskhovrebova, E., Zufferey, S., & Tribushinina, E. (2022b). *Data* [Data set]. OSF. https://osf.io/5khsr/?view_only=0af14a30bfd94b428afc5611ae74ea2e

Tskhovrebova, E., Zufferey, S., & Tribushinina, E. (2022c). *Materials* [Materials]. OSF. https://osf.io/5khsr/?view_only=0af14a30bfd94b428afc5611ae74ea2e

van Silfhout, G., Evers-Vermeul, J., & Sanders, T. J. M. (2014). Establishing coherence in schoolbook texts: How connectives and layout affect students' text comprehension. *Dutch Journal of Applied Linguistics*, *3*(1), 1–29. https://doi.org/10.1075/dujal.3.1.01sil

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

Wei, Y., Evers-Vermeul, J., Sanders, T., & Mak, W. (2021). The role of connectives and stance markers in the processing of subjective causal relations. *Discourse Processes*, *58*(8), 766–786. https://doi.org/10.1080/0163853X.2021.1893551

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*(3), 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). Emerald Group Publishing. https://discovery.ucl.ac.uk/id/eprint/1332007

Wimmer, L., & Ferguson, H. J. (2022). Testing the validity of a self-report scale, author recognition test, and book counting as measures of lifetime exposure to print fiction. *Behavior Research Methods*, *55*(1), 103–134. https://doi.org/10.3758/s13428-021-01784-2

Yefremova, T. F. (Ed). (2000). *Novyy slovar' russkogo yazyka. Tolkovo-slovoobrazovatel'nyy* [New dictionary of the Russian language. With definitions and derivations]. Russkiy yazyk.

Yevgen'eva, A. P. (1999). *Slovar' russkogo yazyka: V 4-kh t., RAN, In-t lingvistich. issledovaniy* [Dictionary of the Russian language: In 4 volumes, Russian academy of sciences, Institute of linguistics] (4th ed.). Russkiy yazyk, Poligrafresursy.

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*, 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. https://doi.org/10.1177/0267658315573349