DISCOURSE PROCESSES
2024, VOL. 61, NO. 8, 381–403
https://doi.org/10.1080/0163853X.2024.2325262

Routledge
Taylor & Francis Group

Connective Comprehension in Adults: The Influence of Lexical Transparency, Frequency, and Individual Differences

Merel Scholman[a,b], Marian Marchal[a], and Vera Demberg[c]

[a]Department of Language Science and Technology, Saarland University, Saarbrücken, Germany; [b]Utrecht University, Institute for Language Sciences, Utrecht, the Netherlands; [c]Department of Language Science and Technology, Department of Computer Science, Saarland University, Saarbrücken, Germany

ABSTRACT

The comprehension of connectives is crucial for understanding the discourse relations that make up a text. We studied connective comprehension in English to investigate whether adult comprehenders acquire the meaning and intended use of connectives to a similar extent and how connective features and individual differences impact connective comprehension. A coherence judgment study indicated that differences in how well people comprehend connectives depend on the lexical transparency but not on the frequency of the connective. Furthermore, individual variation between participants can be explained by their vocabulary size, nonverbal IQ, and cognitive reasoning style. Print exposure was not found to be relevant. These findings provide further insight into the factors that influence discourse processing and highlight the need to consider individual differences in discourse comprehension research as well as the need to examine a wider range of connectives in empirical studies of discourse markers.

Introduction

An important part of creating a coherent mental representation of a text is the construction of discourse relations (Hobbs, 1979; Sanders et al., 1992). Discourse relations are logical links between parts of the text, such as cause, contrast, or conjunction. They can be expressed explicitly using connectives such as *because, whereas*, and *in addition*. Connectives provide readers with "processing instructions" on how to relate incoming text input to previously read segments (Britton, 1994). There is a large body of research showing that comprehenders do indeed benefit from the information provided by connectives (see, e.g., Canestrelli et al., 2013; Cozijn et al., 2011; K¨ohne-Fuetterer et al., 2021; Xiang & Kuperberg, 2015). Readers tend to comprehend texts with connectives better (e.g., Kleijn et al., 2019; Millis & Just, 1994) and process the clauses following the connective faster (e.g., Sanders & Noordman, 2000; Van Silfhout et al., 2015).

Many of these studies have focused on relatively frequent or unambiguous connectives, such as *therefore* or *although*. However, the relationship between the connective and the relation sense is not always straightforward or easy to infer, in part because the connectives' meanings are abstract and can be difficult to define (Crosson & Lesaux, 2013a). This raises the question of whether readers are able to accurately understand the meaning of connectives. Especially when connectives are infrequent, occur mainly in a specific domain, or have no information encoded in their lexical form (i.e., when they are not lexically transparent), it may be difficult for readers to infer their meaning. Recent work indicates that adult comprehenders do not perform at ceiling on connective comprehension tasks using French

CONTACT Merel Scholman m.c.j.scholman@coli.uni-saarland.de Department of Language Science and Technology, Saarland University, Campus C7 2, 66123 Saarbrücken, Germany

connectives typical for the written domain (Tskhovrebova, Zufferey, & Gygax 2022; Zufferey & Gygax, 2020a). These studies suggest that less frequent connectives are more difficult to comprehend, likely because comprehenders are not exposed to them enough to acquire their meaning and usage. However, these studies focused on four connectives, specifically selected for their diverging connective features (e.g., frequent versus infrequent; polyfunctional versus monofunctional). It is unclear whether these results generalize to other connectives, and what the influence is of lexical transparency on connective comprehension.

Further, it is likely that not all comprehenders are equally proficient at understanding connectives. Relevant factors that can play a role in comprehension include linguistic experience (increased exposure to language equals increased opportunity to acquire connectives) and general reasoning (higher reasoning skills facilitate understanding the high-level textual relations that connectives signal). Indeed, studies indicate that there is individual variation in connective comprehension. Zufferey and Gygax (2020b) report that print exposure (related to linguistic experience) can partly explain such differences, Tskhovrebova, Zufferey, & Gygax (2022) found in a connective cloze task that educational background also affects connective usage, and Tskhovrebova, Zufferey, and Tribushinina (2022) found that both print exposure and vocabulary size are relevant to connective usage. An open question is whether nonlinguistic factors, such as general reasoning abilities, also affect connective comprehension. The current study is the first to address this.

In this contribution, we study connective comprehension in English from two perspectives: we examine connective-internal factors, focusing on a novel connective feature, and we investigate reader-internal factors, including both linguistic and nonlinguistic individual difference measures. Specifically, we aim to investigate whether differences in the comprehension of the meaning and intended usage of connectives are dependent on frequency and lexical transparency and whether any individual variation between participants can be explained by participants' linguistic experience and general reasoning skills.

In what follows, we first provide the context for why connectives can be considered a special class of vocabulary knowledge and discuss literature on the comprehension of connectives. We then discuss comprehender-internal factors that might influence connective comprehension.

Role of connectives in discourse

Connectives can be considered a special component of vocabulary, different from other types of vocabulary (Crosson & Lesaux, 2013b). This is because the meanings of connectives are procedural, abstract, difficult to define, and difficult to infer from context (Crosson & Lesaux, 2013a). Moreover, their meanings are inherently relational: Connectives express the logical relations between clauses and even across sentences. Connective comprehension thus taps into higher-level text processing.

A large body of research has demonstrated the facilitative effect of connectives on discourse processing and comprehension (see, e.g., Asr & Demberg, 2020; Blumenthal-Dram´e, 2021; Canestrelli et al., 2013; K¨ohne-Fuetterer et al., 2021; Kleijn et al., 2019; Millis & Just, 1994; Murray, 1997; Van Silfhout et al., 2015; Xiang & Kuperberg, 2015; Xu et al., 2018). However, a lot of the work in this area has focused on a small set of frequent connectives and often considered connectives to be a class of words with equal difficulty. But connectives are not all equal; in fact, they vary significantly in various ways, such as frequency in various modes, complexity in terms of the relations that they can mark, and the number of relations they can express. The current study focuses on two specific features of connectives, namely their frequency in natural language and their lexical transparency.

English contains a large range of connectives (Das et al., 2020; Webber et al., 2019), but their frequency of occurrence in natural language differs dramatically. For example, connectives such as *and, but*, and *because* are quite common, tend to be acquired at an early age, and are understood well by most people. However, connectives such as *moreover, conversely*, and *consequently* are much less common. They tend to be restricted to the written domain, to which not all comprehenders are likely

exposed to a similar degree. These less frequent connectives could pose more comprehension difficulty to people who are not be exposed to them enough to acquire their meaning and usage.

Studies have shown that the frequency of connectives in natural language is a good predictor of connective comprehension. For example, Nippold et al. (1992) studied comprehension of 10 English adverbial connectives[1] in subjects aged 12 to 23 years in a connective cloze task and sentence completion study and found that comprehension was higher for frequently used connectives. In a series of connective insertion and coherence judgment studies, Tskhovrebova, Zufferey, & Gygax (2022) and Zufferey and Gygax (2020a, 2020b) studied adolescents' and adults' comprehension of four French connectives. Similar to Nippold et al. (1992), the results consistently showed that frequent connectives were mastered better than less frequent ones. By contrast, Wetzel et al. (2020) did not find an effect of frequency in a study of the mastery of 12 French connectives in native French speakers and learners. They hypothesize that this might be because the connectives used in their study were not infrequent enough. It is therefore an open question to what extent the effect of frequency is generalizable to connective comprehension across the board. The current study will evaluate the effect of frequency on comprehension using a larger set of English connectives with various frequencies.

Another factor that likely affects how easily connectives are understood by comprehenders is lexical transparency, which refers to whether a word's composite meaning is encoded in the lexical form of the word itself. Lexically transparent connectives are those for which the meaning can be determined from the components (words or parts of words). Consider the difference between *as a result* and *thus*. The relational meaning of *as a result* as a marker of a result relation is encoded explicitly in the lexical element *result*, and the connective is therefore lexically transparent. In *thus*, the result meaning is not encoded explicitly in the connective's form, and so the connective is lexically opaque.

A body of literature has studied the lexical transparency of compound words (e.g., *strawberry* and *blueberry*; see Auch et al., 2020; Gu¨nther & Marelli, 2019, for an overview). These studies have shown that lexical transparency can influence the ease of processing of compound words (Gu¨nther & Marelli, 2019; Kim et al., 2019; Libben et al., 2003; Marelli & Luzzatti, 2012; Sandra, 1990; Zwitserlood, 1994; but see Juhasz, 2018; Pollatsek & Hy¨on¨a, 2005; Smolka & Libben, 2017). Although connectives differ from compound words in their degree of compositionality, we believe that the same principles and methodologies used in the compound transparency field can be extended to connectives. Wetzel et al. (2020), for example, hypothesize that lexical transparency is the reason for why their two-language participants could understand the French connective *c'est pourquoi* ("that's why"). However, no study has systematically investigated the lexical transparency of connectives.

It should be noted that lexical transparency is not a black-and-white distinction but rather functions on a continuum (Auch et al., 2020). Consider *nevertheless*: It does not have the relational meaning of a concession relation explicitly encoded in its lexical form, but it is also not as opaque as *thus*, because the negation element *never* matches the adversative aspect of the concession relation. Determining the lexical transparency of connectives can therefore be a difficult, subjective task. In the current study, we draw on the literature studying the semantic transparency of compound words to obtain lexical transparency scores in a norming pretest.

Individual differences in connective comprehension

Comprehension of connectives is associated with various potentially influencing factors at the individual level. Previous studies have found variation in connective mastery related to age (Nippold et al., 1992), academic background (Tskhovrebova, Zufferey, & Gygax 2022; van Silfhout et al., 2015; Zufferey & Gygax, 2020b), reading proficiency (Van Silfhout et al., 2015), and linguistic experience (Tskhovrebova, Zufferey, & Gygax 2022; Tskhovrebova, Zufferey, and Tribushinina 2022; Zufferey & Gygax, 2020b). The current study further explores the influence of individual differences by also considering other potential sources of individual differences.

We focus on two linguistic experience-related factors (the individual's general vocabulary knowledge base and print exposure) and two individual-internal factors (nonverbal IQ and cognitive

reasoning style), which relate to general reasoning skills. These components are of particular interest because of their established links to text comprehension (e.g., Swanson et al., 2006; Cipielewski & Stanovich, 1992; Mol & Bus, 2011; Scholz & Scheer, 2020). For example, Freed et al. (2017) studied which reader characteristics are central to text comprehension. They measured participants' linguistic experience and general reasoning skills as well as decoding skills, fluency, working memory, inhibition, and perceptual speed. Their results showed that linguistic experience and general reasoning were the only constructs to affect offline text comprehension; the other constructs did not account for additional variance.

Unlike general measures of text comprehension, the current measure of connective comprehension focuses on more fine-grained word-level knowledge that influences overall text comprehension. While prior work has shown that linguistic experience is related to connective comprehension, the current study is the first to also consider the link between general reasoning and connective comprehension. By looking at both of these factors, we are able to specify the cognitive and linguistic processes that underlie comprehension.

Linguistic experience

Linguistic experience comprises exposure to both spoken and written language. Relating to connective comprehension, two specific constructs that are part of linguistic experience are likely to affect connective comprehension. First, we consider *vocabulary knowledge*: It is likely that a person who has acquired a larger number of words also has a greater probability of acquiring more connectives. We use P. Nation and Beglar (2007)'s Vocabulary Size Test (VST) as a measure of vocabulary knowledge. This test is based on word frequencies in spoken vocabulary.

However, there are two reasons for why vocabulary knowledge might not be the strongest predictor of connective comprehension. First, connectives form a specific component of vocabulary knowledge: knowledge of connectives was found to be a significant predictor of reading comprehension, above and beyond breadth of vocabulary knowledge (Crosson & Lesaux, 2013b). Although vocabulary knowledge and connective knowledge might be correlated, they do not necessarily depend on each other directly. Second, general word knowledge is influenced by spoken as well as written language input, whereas many connectives are restricted to the written domain. It is therefore possible that a measure specifically targeting written language experience would be a stronger predictor.

To study experience with written language specifically, we include a test for *print exposure* (i.e., the amount of written text people have been exposed to). Print exposure has often been used as an approximate for linguistic experience. For example, various studies have shown that print exposure can account for differences in reading comprehension (e.g., Cipielewski & Stanovich, 1992; Mol & Bus, 2011). With respect to discourse marking, Wetzel et al. (2020) and Zufferey and Gygax (2020b) have shown that print exposure can predict connective comprehension. Scholman et al. (2020) found that print exposure can explain variability in comprehenders' sensitivity to alternative signals of discourse relations.

Similar to the studies discussed above, we measure print exposure using the Author Recognition Test (ART, Acheson et al., 2008; Stanovich & West, 1989). For this test, participants are presented with a list of author and nonauthor names and asked to indicate which names they recognize to be authors. The test has been found to be a strong predictor of reading skill, likely because author knowledge is often acquired through reading or other forms of print exposure.

Previous work has found that print exposure is correlated to general vocabulary knowledge (e.g., Cunningham & Stanovich, 1991; Martin-Chang & Gould, 2008; Vermeiren et al., 2023). However, given that the ART targets exposure to written language whereas the VST measures vocabulary knowledge based on spoken data, their respective contributions to connective comprehension may differ.

General reasoning

Compared to linguistic experience, the link between general reasoning and connective comprehension is less well established. This construct is nevertheless interesting to examine because it allows us to tease apart the effect of intelligence from linguistic processing skills. General reasoning involves the ability to reason and think flexibly and underlies people's capability to acquire knowledge and solve problems. It affects the ability to detect and establish meaningful relations between objects, ideas, or situations in general, also known as relational reasoning skills (see Holyoak, 2013). These skills are known to vary between individuals, depending on a person's age, domain knowledge, and familiarity with the reasoning context (Alexander et al., 2016).

In dual process theories of information processing, reasoning is characterized by two processing modes, which respond to information in different ways (Evans, 2011; Evans & Stanovich, 2013; Frankish, 2010; Thompson, 2009). A first, intuitive type of thinking (commonly referred to as System 1) triggers fast, automatic processing based on heuristics and intuitions, whereas a second, more deliberate type (System 2) demands more controlled utilization of limited executive resources. Dual process theories assume that people by default engage in System 1 thinking (cf. the Good-Enough Processing Theory, Ferreira et al., 2002; Ferreira & Patson, 2007). Because System 2 thinking is resource demanding, the extent to which an individual engages in controlled, analytical operations will depend on their cognitive capacities. Indeed, previous studies have shown that individual differences in cognitive abilities are related to performance on tasks that rely on System 2 abilities (e.g., see Barrett et al., 2004; Feeney, 2007; Stanovich, 1999; Stanovich & West, 2008).

System 2 is particularly relevant for complex relational reasoning tasks. Analytical reasoning involves deliberate thinking, rule-based processing, and logical deductions. When faced with relational reasoning problems that require integration of multiple arguments, people are more likely to engage a more deliberate processing style. System 2 can analyze and manipulate relational structures, detect patterns, and draw logical inferences to arrive at accurate interpretations. Following from dual process theories, one might therefore expect that comprehenders who are more prone to engage in deliberate and analytical thinking might show better comprehension of connectives, since these require processing of the relational arguments and their integration with the procedural meaning that the connective expresses.

No work has studied the link between general reasoning and connective comprehension, but prior work has shown a link between general reasoning and general language comprehension. Most studies in this area have focused on children and low-literacy adults (e.g., Swanson et al., 2006; Scholz & Scheer, 2020; Tiu et al., 2003). An exception to this is the aforementioned study conducted by Freed et al. (2017), who found that general reasoning was the only construct, besides linguistic experience, to affect offline comprehension of texts in proficient adults. In their study, they included the Raven's Progressive Matrices Test (RPMT) as a measure of general reasoning ability.

We here measure general reasoning using two constructs. The first construct related to general reasoning that we consider is nonverbal, or fluid, IQ, which can be seen as relating to the capacity a comprehender has to successfully carry out System 2 processing. Nonverbal IQ corresponds to the ability to think abstractly, reason quickly, and solve problems independent of any previously acquired knowledge. To measure nonverbal IQ, we use a shortened version of RPMT (Bilker et al., 2012; Raven & Court, 1938). The test consists of a series of puzzles for which a part is missing. Participants are asked to choose the missing part out of six to eight options. The test therefore inherently includes a relational component: Comprehenders need to detect the relationship between the various pieces of the puzzle to understand which piece is missing. If it is indeed the case connective comprehension requires the ability to infer relations and generate inferences, the RPMT should be a robust predictor of individual differences in connective comprehension.

The second construct related to general reasoning is cognitive reasoning style, which can be seen as the mere willingness to engage System 2 processing. This reflects the way individuals think, perceive information, and engage in deeper reflection. We measured cognitive reasoning style using the Cognitive Reflection Test (CRT, Frederick, 2005). This test consists of questions that

have an obvious, intuitive answer that springs to mind but is in fact incorrect. To answer correctly, participants must inhibit the tendency to respond with the incorrect answer and think more analytically. The CRT is argued to be related to various constructs, such as cognitive capacity, inhibitory control, and cognitive style (Frederick, 2005), but recent work indicates that it is primarily a cognitive measure, strongly linked to intelligence (Otero et al., 2022; Welsh, 2022). We expect readers with higher cognitive reflection skills to evaluate connective usage in our task more critically and therefore to show higher connective comprehension abilities. Such an effect should show specifically for incoherent items, where the connective does not match the relation sense expressed by the relational arguments, since critical evaluation is needed to be able to perform well on these items.

Current study

The current study investigates what factors influence connective comprehension in adults, focusing on connective-internal factors and comprehender-internal factors. Our research questions are as follows: *To what extent can connective frequency and lexical transparency explain how well adults can comprehend connectives? And to what extent do linguistic experience and general reasoning skills influence connective comprehension in adults?* We examine the influence of each subconstruct within linguistic experience (frequency and lexical transparency) and general reasoning (nonverbal IQ and cognitive reasoning style) on connective comprehension to determine their respective contributions.

No prior work has studied the lexical transparency of connectives. We therefore conducted a norming pretest to obtain lexical transparency estimates. The norming pretest design draws from the literature of the semantic transparency of compound words (e.g., *strawberry* versus *blueberry*) by asking participants to rate how predictable the meaning of a word is based on the word's components to obtain transparency estimates (see, e.g., Auch et al., 2020; Gagné et al., 2019). Frequency estimates were obtained using the Corpus of Contemporary American English (COCA, Davies, 2008), which comprises different spoken and written subcorpora.

Connective comprehension will be measured through a coherence judgment task. We include a broader range of connectives compared to previous studies using adult participants (Nippold et al., 1992; Tskhovrebova, Zufferey, & Gygax 2022; Zufferey & Gygax, 2020a, 2020b). These prior studies have investigated 4 French and 10 English connectives[2] that were selected for their dichotomous values on frequency and ambiguity parameters; we here included 20 connectives that vary in frequency.

Each connective was presented twice to participants: once in a coherent connective-relation mapping and once in an incoherent connective-relation mapping. This was done to test whether participants are able to accept correct uses and reject incorrect uses. If comprehenders are less familiar with the exact meaning of a connective, their comprehension of correct and incorrect usages could go in two directions. On the one hand, a participant might be likely to accept an unknown connective's usage in any context, giving a relatively high rating in the coherent condition, but also in the incoherent condition. This would mean that they would score lower in the incoherent condition but still perform well in the coherent condition. Alternatively, a participant might not give high ratings to pairings at all when they are unsure of the connective's meaning. In this case, they would show poorer performance in the coherent condition specifically. Hence, it is possible that the effect of "not comprehending" only shows up in one of the two conditions. Therefore, we take into account that condition might interact with frequency and lexical transparency, which can influence how difficult or comprehensible a connective is to a language user.

The rating strategies that participants can take might also relate to individual differences: If participants with lower general reasoning skills or less linguistic experience have more difficulty with comprehending certain connectives, this might only show up in one particular condition. We therefore take into account the interaction between the individual differences measures and condition.

We specifically expect an interaction between condition and CRT: participants with better CRT skills are expected to be better at evaluating the incorrect usage of connectives than participants with poorer CRT skills, since high CRT participants might be more critical of coherence pairings and thus more sensitive to a mismatch. Given that RPMT also measures general reasoning skills, we also expect a similar interaction between condition and RPMT. We are agnostic with respect to condition-specific effects for linguistic experience: the effect might show up in either of the two conditions (i.e., interaction) or both (i.e., main effect).

Methods

This section details the various tests and measures that were included as well as the analysis procedure. All materials, data, and code are available in an online repository.[3]

Participants

A total of 239 native English speakers (age range 18–69 years; mean age 31 years; 171 women), registered as participants on Prolific, took part in this study. Data from an additional 30 participants were excluded due to low quality, as detailed in below. All participants were born in the United States and were currently living there. Participants from varied educational backgrounds were recruited to ensure heterogeneity in the participant sample: 79 participants had completed postsecondary-level education (an undergraduate degree or higher), 84 participants had completed high school and were students, 76 participants had completed high school or had no formal qualifications and were not students. Participants took part in all tests in one session. Total study duration was 40 minutes on average and participants were reimbursed £5.50.

Materials

Twenty connectives were included in this study. These connectives all connect two main clauses and express a variety of relation types, such as causal, additive, temporal and adversative, contrast, similarity, precedence, and arg2-as-detail. The connectives were perceived as relatively difficult by the authors and in a difficulty norming study.

For each connective, a coherent relation (see Example 1) and incoherent relation (Example 2) was created for the coherent judgment test. All items consisted of two clauses each and presented neutral content, so that background knowledge was not needed to be able to construct the coherence relations. They were based in part on French items created by Wetzel et al. (2020). Items were pretested in a connective insertion test to ensure that they clearly conveyed the intended relation sense.

(1) *Coherent pairing*: Lucy is feeling tired. Even so, she is going to a party.
(2) *Incoherent pairing*: Fran put on her shoes. Even so, she tied her shoelaces.

Participants completed two steps for every item. In the first step, participants saw the two relational arguments without a connective and were asked to provide a connective. This step was meant to encourage participants to infer the intended meaning of the relation based on the two relational arguments (i.e., to mitigate the effect of accommodation). In the second step, participants were presented with the same item but with a connective present and were asked to rate on a scale of 1 to 5 "how well does this linking word fit the sentence?" Four relatively easy filler connectives were included (*before, by contrast, like*, and *specifically*); two occurred three times in the incoherent condition, and two occurred three times in the coherent condition.

Individual difference tests

Participants took part in four individual tests as well as the coherence judgment test in one session in the following order: CRT, ART, the coherence judgment test, RPMT, and VST.

Author recognition test

An automated version of the ART (Acheson et al., 2008) was used as a measure of print exposure. Participants were presented with a list of 130 potential authors names; 65 were real author names from Acheson et al. (2008) and 65 were foils (nonauthor names) from Martin-Chang and Gould (2008).

The names were presented one at a time in alphabetical order by last name. Participants were given 10 seconds to decide whether a name was an author name. They were instructed not to guess and only select names that they were absolutely certain to be author names, since their score would be penalized for falsely identifying foils as authors.

Print-exposure scores were calculated by subtracting the total number of foils that were falsely identified from the total number of authors that were correctly identified. Participants who selected more than 50% of the foils or participants who timed-out on more than 33% of all trials were excluded from the ART analysis.[4]

Vocabulary size test

Receptive vocabulary was tested using a shortened version of the 2,000-word-family VST (adapted from P. Nation & Beglar, 2007). The original test systematically samples 20 word frequency levels established using data from the spoken part of the British National Corpus (BNC Consortium, 2007) and the COCA (Davies, 2008). Each frequency level comprises the 1,000 most frequent "word families" in ascending order of difficulty. The full test contains 5 items for each vocabulary level and 100 items in total. Since participants in this study had to complete four other tasks, the test was abridged to avoid fatigue. Levels 1 to 4, which contain the most frequent words (e.g., *time, circle*), were omitted, as well as levels 19 and 20, which contain the most difficult words (e.g., *casuist, sylvan*). The abridged version contained 70 items.

Each test word was presented in a short nondefining context and followed by 4 definitions; the participant's task was to choose the best definition for the target word. The participants' scores corresponded to the number of correctly answered items.

Raven's progressive matrices test

A 12-item version of RPMT (adapted from Bilker et al., 2012) was included as a measure of participants' nonverbal intelligence. Bilker et al. (2012) found that the score on their abbreviated test of nine items correlated almost perfectly with the full-length Standard Progressive Matrices Test. We used Bilker et al. (2012)'s nine-item version and expanded it with two additional "easy" items from the Standard Progressive Matrices Test (one from level A and one from B) and the final item from the Advanced Progressive Matrices Test, to account for possible floor and ceiling effects, respectively.[5]

The test consisted of a series of images with a pattern for which a part is missing. Participants were asked to choose the missing part out of six to eight answer options. The task increased in difficulty level. The participants saw a timer on their screen. They were told that their time was not limited but that the task should take about 5 minutes and that they should avoid thinking too long. The score corresponded to the number of correctly solved items.

Cognitive reflection test

A 10-item version of the CRT (based on Frederick, 2005) was included as a measure of their cognitive processing skills. The test contained six critical questions—three verbal and three involving numeracy—for which there was an "obvious" or intuitive answer which was incorrect. The test also contained four nontrick decoy questions to prevent the participants from expecting to be tricked every time.

Since CRT is known to be affected by familiarity (Woike, 2019), we used an updated CRT version. Our version consisted of questions from previously proposed versions of CRT (Baron et al., 2015; Primi et al., 2016; Sirota & Juanchich, 2018; Thomson & Oppenheimer, 2016; Toplak et al., 2014). For each question, participants were asked to indicate if they had seen this question before (13% of observations were affected; 7 participants had seen more than two questions before).

The score corresponds to the proportion of correctly answered questions. If the participants reported having seen one or two of the critical questions before, we removed their answers to those questions from consideration and computed an adjusted score out of five or four questions, respectively. Participants who had seen three or more critical questions before were excluded from analysis for this test based on the assumption that three out of six questions is not enough to reliably estimate the CRT score.

Connective measures

The current study tests the influence of a connective's frequency and lexical transparency on comprehension. This subsection explains how these estimates were obtained.

Connective frequency estimates

Some connectives in our study can also occur in a nonconnective usage (e.g., *for* as a preposition rather than a causal connective). Raw frequency measures therefore likely do not reflect their true usage.[6] We therefore estimated a connective's frequency based on their occurrence in the COCA (Davies, 2008), which is a 1.1 billion word corpus of American English and comprises different spoken and written subcorpora.

We randomly extracted 50 occurrences of each connectives and counted the proportion of connective usages. If any of these 50 instances were nonconnective usages, we annotated another 150 instances of the connective for discourse relation usage (this was done for *accordingly, even so, for, given that, hence, indeed*, and *in fact*). We then calculated the proportion of connective usage per connective and used these proportions to estimate number of occurrences per million words in the COCA accordingly. Log-transformed frequency per million words was used as the variable of frequency in the analyses.

Table 1 provides the frequency estimates per connective. The mean frequency estimate for all connectives was 23 occurrences per million words (min = 2, max = 127).

Lexical transparency estimates

To obtain lexical transparency estimates of the included connectives, we conducted a pretest in which 52 native English speakers (age range 20–73; mean age 34; 28 women) from Prolific took part. The average study duration was 6 minutes, and participants were reimbursed £1.50. We recruited participants from varied academic backgrounds because readers likely vary in their familiarity with the

Table 1. Comprehension score per connective, including transparency and frequency per million words. *Mean*: average transformed coherence judgment score per connective.

Connective	Mean	Trans.	Freq.	Connective	Mean	Trans.	Freq.
as if	4.72	87	127	for	4.26	66	18
given that	4.67	82	13	notably	4.21	65	3
as though	4.62	81	29	consequently	4.05	85	8
on the contrary	4.57	83	4	conversely	4.03	61	4
hence	4.48	66	8	albeit	3.93	36	8
indeed	4.46	60	43	furthermore	3.87	73	21
accordingly	4.43	75	4	thereafter	3.80	74	7
even so	4.43	74	7	subsequently	3.72	64	2
nonetheless	4.42	63	18	moreover	3.59	63	29
in fact	4.40	88	89				

components of the connectives (e.g., the component "subsequent" in *subsequently*), which is an important aspect of the connective's lexical transparency.

Participants were presented with the connectives one at a time. For each connective they also saw the components that make up the connective, a definition of the connective, and an example sentence containing the connective. The components were included to highlight the relevant parts of the words. Participants were asked to rate the predictability of the meaning of the connective based on its components ("predictability" rather than "transparency" was chosen, cf., Gagné et al., 2016, 2019; Libben et al., 2003) using a scale ranging from 0 to 100. They were instructed that 0 indicates that the meaning of the connective is not predictable at all from the connective itself and that 100 indicates that the meaning of the connective is very predictable from the connective itself.

We conducted a split-half reliability assessment for the transparency ratings. The Spearman-Brown corrected split-half reliability was high: $r = .82$. We used the raw ratings provided by the participants to calculate median lexical transparency estimates per connective, which are provided in Table 1. The median transparency rating for all connectives was 69 of 100 (mean of all median ratings = 70.6, min = 37, max = 88). A boxplot showing the distribution of the ratings per connective is presented in the Appendix. Note that the transparency and frequency scores were not correlated significantly with each other ($r = .29$, $p = .22$).

Analysis procedure

Data exclusion

Crowd-sourced participants have been used successfully in a variety of discourse-related tasks, including coherence judgment studies (e.g., Asr & Demberg, 2020; Zufferey & Gygax, 2020b) as well as individual difference tests (e.g., Crump et al., 2013; Scholman et al., 2020; Zufferey & Gygax, 2020b). The main concern with using these types of paradigms is that participants are more likely to not take part seriously (i.e., providing random answers or cheating in another way) in an uncontrolled setting like in crowd-sourcing compared to a lab-based setting. This is especially likely to occur in tasks that make it easy to provide random answers, such as with the ART task, in comparison to tasks which require written responses. To control for this, we implemented quality checks for the ART used in this study, as elaborated on in the ART description. We also excluded data for the CRT based on participants' prior exposure to this test, as elaborated on in the CRT description. We here describe the exclusions that resulted from these checks.

Data from 30 participants were removed to ensure data quality. Of these 30 participants, 8 participants were excluded because they selected more than 50% of the foils on the ART (3% of all ART submissions). Fifteen participants were excluded because they timed out on more than 33% of all trials on the ART (5.6% of all ART submissions). Finally, data from seven participants were excluded because they indicated they had seen more than two CRT questions before (2.6% of all CRT submissions). All data from these participants were excluded from the analyses.

We furthermore removed observations for one connective (*insofar as*) from the coherence judgment test because the item was paired with an incorrect connective in the coherent condition; the arguments of the item elicited a result inference, whereas the connective expressed a reason relation. The remaining dataset contained 9,082 experimental observations (239 participants × 19 connectives × 2 conditions).[7]

Data analysis

To answer the question of how frequency, lexical transparency, and the four individual difference measures affect connective comprehension, we studied the influence of these factors on participants' performance on the connective comprehension task. In parity with a prior connective comprehension study (Zufferey & Gygax, 2020b), the response variable used in the analyses was the Likert responses (scale of 1–5), with scores on the incoherent condition inverted such that a score of 5 always reflects accurate recognition of either a coherent or incoherent pairing.[8] The fixed effect of condition was deviation-coded (−.5 = coherent; .5 = incoherent), frequency was log-transformed and centered, and the fixed effects of lexical transparency and IDs were centered.

Results were modeled using ordinal mixed-effect regression models. Ordered logit models were created with the ordinal::clmm function and evaluated using the lme4 package within the statistical software R (Bates, 2023; Christensen, 2019; R Development Core Team, 2008). Models were fit with flexible, nonequidistant thresholds. We used a step-down approach to model selection, starting with the fixed effects for condition, connective features (frequency and transparency) and individual differences (ART, VST, RPMT, and CRT). The full model also contained interactions between condition and fixed effects of the connective features as well as between condition and CRT to test whether effects of the connective and CRT skills are modulated by the coherent or incoherent usage of the connective in its context. In case of interactions between fixed effects in the final model, we interpret the direction of the interaction using model estimates. Upon selection of the final model, we added maximal random effect structures (Barr et al., 2013); the final converging maximal random effect structure is reported together with the model.

Results

Table 1 presents the inversed Likert ratings per connective (averaged over condition); higher ratings indicate that comprehenders more accurately rated the coherence of the connective-relation pairing. The results show that *as if* yields the highest performance, followed by *as though* and *given that*. *Moreover* yielded poorest performance, followed by *subsequently* and *thereafter*.

In what follows, we first provide descriptive statistics relating to the individual difference measures. Next, we present the analyses regarding the effects of connective features and individual reader characteristics on the comprehension scores.

Individual difference measures: descriptive statistics

Table 2 presents the descriptive statistics and reliability estimates for each test. Reliability was estimated using the split-half correlation (odd items versus even items), corrected for length with the Spearman-Brown formula; Cronbach's alpha is also reported. The estimates show good reliability for all tests ($r > .7$), with the exception of the CRT ($r = .67$), which shows reliability estimates comparable to what was previously found in the literature (Stieger & Reips, 2016).[9]

Table 3 presents the correlations among the individual difference measures and the connective comprehension score. As expected, ART and VST, both measures of linguistic experience, strongly correlate with each other. The correlation between RPMT and CRT is also moderate and in line with

Table 2. Descriptive statistics for the included tests. *Kurt.*: kurtosis; *Rel.*: Spearman-Brown corrected split-half reliability; *α*: Cronbach's alpha.

Test	Possible range	Observed range	Mean	Median	SD	Skew	Kurt.	Rel.	*α*
ART	0–65	0–60	19.70	18.00	11.70	0.96	3.95	.91	.89
VST	0–1	0.19–1	0.74	0.74	0.11	−1.09	6.38	.86	.86
RPMT	0–12	2–12	7.72	8.00	2.15	−0.43	2.68	.72	.70
CRT	0–1	0–1	0.31	0.20	0.25	0.76	2.92	.67	.61
Connective comp.	1–5	2.92–4.89	4.25	4.34	0.44	−0.94	3.31	.90	.86

Table 3. Pearson's correlations for individual difference measures and the connective comprehension score, corrected for multiple comparisons. *Correlation significant at the $p < .01$ level.

	ART	VST	RPMT	CRT
VST	.61*	–		
RPMT	.21*	.44*	–	
CRT	.29*	.33*	.37*	–
Connective comp.	.34*	.64*	.43*	.23*

previous results (e.g., Meyer et al., 2018; Primi et al., 2016). We further note that performance on the connective comprehension test is correlated with the VST and RPMT, indicating that participants who score higher on these tests also tend to score higher on the connective comprehension test.

Performance on the coherence judgment task

To evaluate the validity of the ratings provided in the second step of the task, we also considered insertions in the first step, for which participants were asked to freely insert a connective to express the logical relation between the two clauses. Specifically, we assessed whether participants did not infer the relation expressed by the coherent connective or erroneously inferred the relation expressed by the incoherent connective.[10] This procedure revealed that for 92% of the data, the inferred relation in Step 1 matched Step 2′s connective sense in the coherent and mismatched the connective sense in the incoherent condition. Note that repeating the analysis below with this subset of the data shows qualitatively similar results.

As expected, participants assigned lower Likert ratings to connectives that occurred in incoherent pairings (average rating of 1.80 on a scale of 1–5) compared to those that occurred in coherent pairings (average rating of 4.27). In the remainder of the analyses, we use only the inverted ratings as the response variable. This inverted score shows that there was little difference in how accurately readers can recognize coherent versus incoherent usage of connectives: items in the coherent condition received a mean score of 4.27, whereas items in the incoherent condition received a mean score of 4.20.

Table 2 shows that there is variability between participants in their performance on the coherence judgment test, with the top scoring participant achieving a mean score of 4.9 (out of 5) and the lowest scoring participant obtaining a mean score of 2.9. Likewise, there was strong variability in the comprehension of connectives as shown in Table 1.

The main question in the current study is whether such variability in connective comprehension can be explained by connective factors and individual reader characteristics. To explore this, we modeled the transformed response variable in an ordinal mixed-effect regression model. The full and final models are presented in Table 4.[11] We first discuss results regarding the connective factors and then focus on the results regarding the individual differences.

Connective-internal factors influencing connective comprehension

Table 4 shows an interaction between condition and transparency. This effect is visualized in Figure 1: In the coherent condition, connectives with higher transparency scores are more likely to receive higher judgments than connectives with lower transparency scores, as indicated by the lighter colors being more dominant when the transparency estimate increases. In contrast, Figure 1 shows a negligible effect of transparency in the incoherent condition. Frequency was not included in the final model as it did not significantly improve model fit.

Comprehender-internal factors influencing connective comprehension

In relation to the linguistic experience subconstructs, the final model showed a significant main effect of VST: Participants with higher vocabulary size also perform better on the connective comprehension test. The effect of ART was not significant when VST was included in the model.

In relation to the general reasoning subconstructs, the final model showed a significant main effect of RPMT: Participants with a higher score on the IQ test also perform better on the connective comprehension test. The main effect of CRT was not significant, but the results did show an interaction between CRT and condition, indicating that the effect of CRT is different in the two conditions. The model estimates of the interaction effect of CRT and condition are visualized in Figure 1. In the coherent condition, the lighter colors are less dominant as CRT increases. This means that participants with higher CRT scores were more likely to give these coherent item-connective pairings lower ratings than participants with lower CRT scores. By contrast, in the incoherent condition, the lighter colors are more dominant as CRT increases.

Table 4. Regression coefficients and test statistics from the full and final ordinal mixed-effects model.

	Full model				Final model			
	β	SE	z	*p*	β	SE	z	*p*
Fixed effects								
Condition	−0.10	0.26	−0.37	.71	−0.10	0.26	−0.37	.71
Lexical transparency	1.43	0.70	2.05	**.04**	1.39	0.48	2.88	<.**005**
Frequency	0.87	0.61	1.44	.15				
ART	0.10	0.38	0.27	.79				
VST	4.21	0.58	7.25	<.**001**	4.40	0.60	7.29	<.**001**
RPMT	0.74	0.31	2.36	**.02**	0.75	0.34	2.22	**.03**
CRT	−0.15	0.25	−0.60	.55	−0.12	0.27	−0.46	.64
Condition: Lexical transp.	−2.96	1.04	−2.86	<.**01**	−2.94	0.99	−2.97	<.**001**
Condition: Frequency	0.69	0.90	0.76	.45				
Condition: ART	0.41	0.77	0.54	.59				
Condition: VST	−0.14	1.16	−0.12	.90				
Condition: RPMT	0.51	0.63	0.81	.42				
Condition: CRT	0.80	0.51	1.56	.12	1.10	0.47	2.35	**.02**
Random effects								
σ2	3.29				3.29			
τ00 ID	.64				.69			
τ00 Class	.46				.58			
ICC	.38				.42			
Observations	9,082				9,082			
Marg. R^2 /Cond. R^2 =	.130/.459				.112/.483			

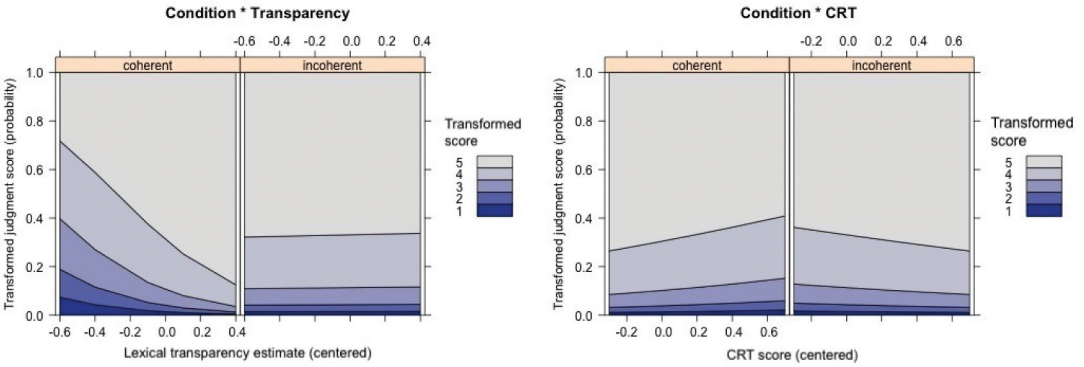

Figure 1. Effect display for the interaction effects between condition and lexical transparency (left) and condition and CRT (right) in the final model.

Note that we analyze the transformed scores here, with higher ratings in the incoherent condition reflecting lower raw Likert ratings. Hence, the interaction between CRT and condition seems to be driven by the fact that participant with higher CRT scores tend to give lower ratings in general. Indeed, an analysis using the raw (non-inverted) scores shows a main negative effect of CRT (β = −0.48, SE = 0.23, z = −2.05, p = .04). This is addressed further in the discussion.

Discussion

We set out to investigate what factors influence connective comprehension. Specifically, we studied whether differences in how well people understand the meaning and intended usage of connectives are dependent on frequency and lexical transparency, and whether any individual variation between participants can be explained by participants' print exposure, vocabulary size, nonverbal IQ, and cognitive processing style. The results showed that people do in fact differ in how well they understand the meaning and intended usage of English connectives and that this

ability is dependent on the connectives' lexical transparency as well as the comprehenders' linguistic experience and general reasoning skills.

Connective-internal features

The results showed that not all connectives are understood to the same extent, with *moreover* being the poorest-scoring connective and *as if* being the best-scoring connective. Connective comprehension can in part be explained by one specific connective-internal factor: more transparent connectives yield better performance in the coherent condition. This indicates that when readers infer the intended discourse relation sense for two segments, they then use this information to confirm this reading. However, for incoherent connective-relation pairings, readers do not use the information provided by the connectives to evaluate the pairing as incoherent. Rather, they reinterpret the relation to accommodate the connective. Consider the following example: *Jack had his passport photo taken. . . . he applied for a new passport*. Typically, people don't get a new passport because they had their photo taken; instead, people get their photo taken because they are getting a new passport. Indeed, when looking at the insertions in Step 1 for this particular item, it seems that most participants inferred a reason relation when presented with only the relational arguments (only 9% of participants incorrectly inferred a result relation in Step 1). However, when presented with the same item including the result connective *consequently*, many participants rated it rather coherent. Hence, participants accommodated the connective despite arriving at a different interpretation in the absence of a connective, which suggests that people are very reliant on the information provided by the connective.

We also tested for the influence of connective frequency on connective comprehension, but the results showed no effect of frequency. This converges with Wetzel et al. (2020), who also did not find an effect of frequency. However, it contradicts findings from previous studies (e.g., Nippold et al., 1992; Tskhovrebova, Zufferey, & Gygax 2022; Zufferey & Gygax, 2020a, 2020b). One possible factor explaining this is that the prior studies that found an effect of frequency selected the connectives to be on extreme ends of the scale (very frequent versus very infrequent), whereas we selected connectives based on other parameters and therefore did not dichotomize frequency as such.

Finally, we reflect on connective-specific results. The finding that *moreover* yielded the poorest performance is interesting, given that it has also been discussed in previous connective comprehension studies as a connective yielding distinctly poor results. Nippold et al. (1992) and Goldman and Murray (1992) both report that *moreover* yielded particularly low scores in their comprehension studies; Goldman and Murray (1992)'s participants also indicated that they did not know what *moreover* meant. Furthermore, our intuition was that *moreover* is a transparent connective ("more" indicating "in addition"); however, it was rated as relatively opaque. The current study thus provides further indications that comprehenders seem to struggle with this connective.

Other connectives that appeared to be difficult to comprehend for participants, in parity with results from Nippold et al. (1992), include *consequently* and *conversely*. However, performance on certain connectives did not converge with earlier findings. For example, Goldman and Murray (1992) reported that *nonetheless* was also considered particularly difficult by their participants, but in our study, participants performed relatively well on this connective. Nippold et al. (1992) found that participants performed quite well on the connective *furthermore*, while in our study, participants did not perform well on this connective. These differing results could be due to differences in the nature of the employed tasks (Nippold et al. (1992) and Goldman and Murray (1992) both used multiple-choice tasks rather than a rating task) or possible differences in the items, but further research is needed to confirm this.

Comprehender-internal features

The current study also showed that comprehender-internal features can explain part of the variability between readers in how well they understand connectives. In terms of the subconstructs related to linguistic experience, previous work has mainly focused on the relationship between connective comprehension and print exposure (but see Tskhovrebova, Zufferey, and Tribushinina 2022). The current study took a more comprehensive approach to operationalizing linguistic experience and found that vocabulary size has a greater effect on connective comprehension than print exposure. Specifically, ART was not found to be predictive of connective comprehension, unlike previous findings (Tskhovrebova, Zufferey, & Gygax 2022; Wetzel et al., 2020; Zufferey & Gygax, 2020b note that they did not consider VST). This is likely due to vocabulary size already explaining much of the variance, given that ART is predictive of comprehension if VST is not included in the model. In other words, while print exposure might be related to connective comprehension, vocabulary size has greater predictive power.[12] This is interesting, given that many of the connectives included in the current study are typical for the written domain and the ART was designed to more directly target exposure to written text than the VST. We recommend future work on individual variation in connective comprehension to include vocabulary size as a variable for linguistic experience.

A critical note has to be made on the applicability of the print exposure measure to the participant sample used in the current study. A large body of research has confirmed the ART's usefulness as a measure of print exposure, but most of this work involved university-level native readers of English. McCarron and Kuperman (2021) studied the ART's reliability in native English speakers with a below-university level of reading proficiency and in non-native readers of English. Their results showed that ART comes with a relatively high standard error of measurement in a nonuniversity educated student sample, indicating that the test is less reliable in samples with lower-educated participants.

Regarding the relationship between general reasoning and connective comprehension, the current study is the first to show that general reasoning does, in fact, play a role: Comprehenders with higher general reasoning skills are also better at understanding the meaning and intended usage of connectives, independent of linguistic experience. We believe this can be attributed to the link between general reasoning and relational reasoning in a broader sense. Indeed, the results showed that the test that taps into relational reasoning skills directly, namely RPMT (which requires participants to recognize patterns and infer relations between separate information units), could predict performance on both the coherent and incoherent pairings.

The CRT, which was also included to measure general reasoning, showed a different effect on connective comprehension compared to the RPMT. Specifically, the effect of cognitive reflection skills differed per condition, with comprehenders with better CRT skills being more likely to provide lower raw Likert ratings to both coherent and incoherent pairings, therefore leading to a poorer inversed score on the coherent condition. This result does not match our prediction that comprehenders with higher cognitive reflection skills evaluate incorrect connective usage more critically, which should lead participants to score particularly well on items in the incoherent condition. Rather, the results suggest that people with higher cognitive reasoning skills are more critical of connective-relation mappings in general than people with lower cognitive reasoning skills. One possible explanation for this effect is the mismatch between the formal connectives and the narrative structure of the items. Participants who are more critical in general might have consistently assigned slightly lower coherence scores on the scale (both in the coherent and the incoherent condition) due to the mismatch, despite the instruction stating that the pairings should not be judged based on how "nice" they sound. Further research is needed to investigate what contributes to this effect of CRT on coherence judgments. A first step in this direction would be to use a binary coherence judgment rather than a scale.

In sum, the results on comprehender-internal factors influencing connective comprehension showed that both linguistic experience and general reasoning are relevant. We note that this research into comprehender-internal factors was not exhaustive; other factors might also be relevant. Specifically, the current study measured vocabulary breadth, but vocabulary depth could also be

relevant: comprehenders differ in how much is known, or the quality of the information, about particular words. Past studies have shown that breadth and depth contribute independently to explaining variance in reading comprehension (Binder et al., 2017; K. Nation & Snowling, 2004; Tran et al., 2020). Further, given that the connective comprehension test required participants to produce a written connective, it could be interesting to include expressive vocabulary tests in future work in this area. Such tests measure an individual's ability to use language actively by generating words and meanings.

Moreover, linguistic abilities that are relevant to text comprehension might be related to connective comprehension as well. This includes skills such as decoding, phonemic awareness, orthographic awareness, and morphological awareness. This latter skill might be particularly relevant in the context of the current study. Morphological awareness is the recognition and understanding of and ability to manipulate the smallest meaningful units in words (see Apel, 2014). It has been found to be strongly associated with vocabulary knowledge, as morphological knowledge can help individuals infer the meanings of unfamiliar words by breaking them down into smaller meaningful parts (see Spencer et al., 2015). Morphology appears to impact comprehension directly as well as indirectly through word reading skills (Perfetti et al., 2005); indeed, empirical studies have shown a link between morphological awareness and reading comprehension in children (e.g., Deacon et al., 2014; Levesque et al., 2017; Tong et al., 2011), adult with low literacy skills (e.g., Fracasso et al., 2016; Tighe & Binder, 2015), and adult proficient readers (Guo et al., 2011; Kotzer et al., 2021). Relating this to the current study, note that most of the included connectives were morphologically complex, and thus require understanding of the bases and suffixes of words. Comprehenders with poorer morphological awareness might be less capable of recognizing, understanding and manipulating the morphological units of the connectives, which could be another explanatory factor of connective comprehension, or of the influence that lexical transparency has on connective comprehension. Future studies will hopefully provide more insight into these issues.

Implications for language comprehension and production

Prior research has shown that connectives facilitate comprehension, which has been attributed to the assumption that comprehenders use connectives as "processing instructions," signaling to readers how to connect relational arguments. However, the facilitative effect of connectives is likely diminished when comprehenders do not understand the meaning of the connective, and this might differ for individual comprehenders. Our study shows that adults do not always understand the meaning provided by the connective. The connectives included in the current study were more typical of the written domain compared to the spoken domain. Exposure to written connectives is more difficult to come by, as this only comes through reading. It could be beneficial for comprehenders to be exposed to more and perhaps more explicit training of connectives typical for the written domain (such as *moreover, albeit*, and *subsequently*) in school curricula. Crosson and Lesaux (2013a) discuss various considerations for connective vocabulary instruction.

An open question is how connective comprehension affects discourse processing. Connectives have been shown to be important signals that readers exploit during online processing, but most studies showing that connectives result in faster online reading times have focused on a limited set of frequent connectives. It would be interesting to study the effects on processing for connectives that yielded poor performance on the comprehension test, such as *moreover* and *albeit*. It is likely that such connectives are only beneficial to the processing of some readers who are able to exploit the connective's information signal. Expanding the research in the field of discourse to less-considered connectives would provide a more comprehensive and generalizable understanding of how connectives impact readers.

We also consider the relationship between connective comprehension and connective production an interesting avenue for future research. Does a larger connective vocabulary knowledge also imply that those comprehenders are more likely to use those connectives in their own speech or writing? In

acquisition literature, it has been shown that production of connectives does not equal accurate comprehension, nor that the connectives are produced in accurate relation-connective mappings (e.g., Cain & Nash, 2011; Cain et al., 2005; Crowhurst, 1987; Geva & Ryan, 1985; Knoepke et al., 2017; Volodina & Weinert, 2020; Welie et al., 2017). It would be interesting to determine whether this extends to adult usage of connectives as well.

In sum, the current study investigated the factors influencing adults' connective comprehension. Being able to establish discourse relations in a text is crucial to creating a mental representation of that text. Given that connectives guide readers in this interpretation, connective comprehension is an essential component of discourse processing. A better understanding of the factors that influence connective comprehension can provide more insight into factors that influence discourse processing, and into possible comprehension difficulties that readers might experience. The results from the current study indeed show that not all connectives are understood equally well (depending on lexical transparency, but not frequency), and that not all comprehenders are able to understand connectives equally well (dependent on both linguistic experience and general reasoning). These results therefore emphasize the need for considering individual differences in discourse comprehension, and for examining a broader range of connectives in studies of discourse markers, for example, to study the influence of more opaque, ambiguous, or difficult connectives on discourse processing.

Notes

1. Connectives included were *consequently, moreover, similarly, furthermore, therefore, nevertheless, however, contrastively, conversely*, and *rather*.
2. But see Goldman and Murray (1992), who included 23 connectives that occurred at least 10 times per million words. However, they aggregated the results into four general categories of connectives, making a direct comparison between their results and ours impossible.
3. https://osf.io/85a3e/
4. The foil selection criterion was established based on the notion that participants who selected such a disproportionate amount of foils did not follow instructions to the task, which explicitly stated they could only select names that they were absolutely certain to be author names. It could be argued that the final ART score for such participants would be corrected for guessing (e.g., a participant selecting 41 author names and 35 foils would receive a score of 41–35 = 6), but such a score would not reflect the same as the score of a participant who was much more conservative in their selection (e.g., selecting 6 authors and 0 foils). The time-out criterion was established because it is difficult to reliably estimate the ART score for participants who timed out on a large number of trials; did they not recognize the names or did they not pay attention during the trial? The 15 participants whose scores were removed based on this criterion timed out on 90 out of 130 trials on average, indicating that they did not participate in this task seriously or simply left the task running without participating.
5. The labels of the items that were included were a8, a11, b7, b12, c4, c12, d7, d12, e1, e5, e7, and ad36.
6. Connective frequency can be obtained from annotated corpora, but these corpora are often limited with respect to modality, domain and size.
7. The experimental design and analysis procedure were preregistered on OSF (https://osf.io/hymgu). A more appropriate analysis based on reviewers' feedback of the data is presented here. The findings are in line with the results when following the preregistered analysis.
8. We also repeated the analyses on a nontransformed score of raw Likert ratings, and the results remained the same, albeit inversed (i.e., a main effect in the model with the transformed response variable appeared as an interaction effect with condition in the model with the raw response variable).
9. This lower reliability might be attributed to the small number of questions included in the test and to the fact that some participants might have seen certain questions previously, which impacts the reliability of the score (see Stieger & Reips, 2016). For this reason, we calculate the CRT score per participant by excluding responses to questions that participants indicate they have seen previously.
10. For the coherent condition, it is important that readers interpret the relation as conveyed by the connective in Step 2. Step 1 insertions that did not match such an interpretation were excluded from this follow-up analysis. Since the free insertions in the first step often contained ambiguous connectives (e.g., "and"), we considered all relation senses that could be conveyed by such connectives. Crucially, for the incoherent condition, readers should *not* have interpreted the relation sense provided by the connective in Step 2. Thus, for the incoherent

condition, we excluded all cases in which a connective was inserted in Step 1 that could convey the relation marked by the infelicitous connective in Step 2.

11. Random effect structure of final model: (1 + Condition + VST + RPMT + CRT | connective) + (1 + Condition × Transparency | subject).
12. Note that Tskhovrebova, Zufferey, and Tribushinina (2022) did find an effect of both ART and VST in a connective cloze task, but this assesses connective usage rather than comprehension.

Acknowledgments

This project received funding from the European Research Council (ERC) under the European Union's Horizon 2020 programme (ERC Starting Grant 'Individualized Interaction in Discourse', grant agreement No. 948878).

Disclosure statement

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

Funding

This work was supported by the European Research Council [948878].

Data availability statement

All materials, data and code are available in an online repository at https://osf.io/85a3e/.

References

Acheson, D. J., Wells, J. B., & MacDonald, M. C. (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

Alexander, P. A., Jablansky, S., Singer, L. M., & Dumas, D. (2016). Relational reasoning: What we know and why it matters. *Policy Insights from the Behavioral and Brain Sciences*, *3*(1), 36–44. https://doi.org/10.1177/2372732215622029

Apel, K. (2014). A comprehensive definition of morphological awareness: Implications for assessment. *Topics in Language Disorders*, *34*(3), 197–209. https://doi.org/10.1097/TLD.0000000000000019

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

Auch, L., Gagné, C. L., & Spalding, T. L. (2020). Conceptualizing semantic transparency: A systematic analysis of semantic transparency measures in english compound words. *Methods in Psychology*, *3*, 100030. https://doi.org/10.1016/j.metip.2020.100030

Baron, J., Scott, S., Fincher, K., & Metz, S. E. (2015). Why does the cognitive reflection test (sometimes) predict utilitarian moral judgment (and other things)? *Journal of Applied Research in Memory and Cognition*, *4*(3), 265–284. https://doi.org/10.1016/j.jarmac.2014.09.003

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Barrett, L. F., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin*, *130*(4), 553. https://doi.org/10.1037/0033-2909.130.4.553

Bates, D. (2023). *Package 'lme4'* (Version 1.1). https://cran.r-project.org/web/packages/lme4/lme4.pdf

Bilker, W. B., Hansen, J. A., Brensinger, C. M., Richard, J., Gur, R. E., & Gur, R. C. (2012). Development of abbreviated nine-item forms of the Raven's standard progressive matrices test. *Assessment*, *19*(3), 354–369. https://doi.org/10.1177/1073191112446655

Binder, K. S., Cote, N. G., Lee, C., Bessette, E., & Vu, H. (2017). Beyond breadth: The contributions of vocabulary depth to reading comprehension among skilled readers. *Journal of Research in Reading*, *40*(3), 333–343. https://doi.org/10.1111/1467-9817.12069

Blumenthal-Dram´e, A. (2021). The online processing of causal and concessive relations: Comparing native speakers of English and *German*. *Discourse Processes*, *58*(7), 642–661. https://doi.org/10.1080/0163853X.2020.1855693

BNC Consortium. (2007). *The British National Corpus*. University of Oxford.

Britton, B. K. (1994). *Understanding expository text: Building mental structures to induce insights*. 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. 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
Christensen, R. H. B. (2019). *Ordinal—regression models for ordinal data* (Version 2019.12-10). http://www.cran.r-project.org/package=ordinal/
Cipielewski, J., & Stanovich, K. E. (1992). Predicting growth in reading ability from children's exposure to print. *Journal of Experimental Child Psychology*, *54*(1), 74–89. https://doi.org/10.1016/0022-0965(92)90018-2
Cozijn, R., Noordman, L. G., & Vonk, W. (2011). Propositional integration and world-knowledge inference: Processes in understanding 'because' sentences. *Discourse Processes*, *48*(7), 475–500. https://doi.org/10.1080/0163853X.2011.594421
Crosson, A. C., & Lesaux, N. K. (2013a). Connectives: Fitting another piece of the vocabulary instruction puzzle. *The Reading Teacher*, *67*(3), 193–200. https://doi.org/10.1002/TRTR.1197
Crosson, A. C., & Lesaux, N. K. (2013b). 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
Crowhurst, M. (1987). Cohesion in argument and narration at three grade levels. *Research in the Teaching of English*, 185–201. https://www.jstor.org/stable/40171109
Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's mechanical Turk as a tool for experimental behavioral research. *PloS one*, *8*(3), e57410. https://doi.org/10.1371/journal.pone.0057410
Cunningham, A. E., & Stanovich, K. E. (1991). Tracking the unique effects of print exposure in children: Associations with vocabulary, general knowledge, and spelling. *Journal of Educational Psychology*, *83*(2), 264. https://doi.org/10.1037/0022-0663.83.2.264
Das, D., Stede, M., Ghosh, S. S., & Chatterjee, L. (2020). DiMLex-Bangla: A lexicon of Bangla discourse connectives. In *Proceedings of the 12th language resources and evaluation conference* (pp. 1097–1102).
Davies, M. (2008). *The corpus of contemporary American English: 450 million words, 1990–present*. http://corpus.byu.edu/coca/
Deacon, S. H., Kieffer, M. J., & Laroche, A. (2014). The relation between morphological awareness and reading comprehension: Evidence from mediation and longitudinal models. *Scientific Studies of Reading*, *18*(6), 432–451. https://doi.org/10.1080/10888438.2014.926907
Evans, J. S. B. (2011). Dual-process theories of reasoning: Contemporary issues and developmental applications. *Developmental Review*, *31*(2–3), 86–102. https://doi.org/10.1016/j.dr.2011.07.007
Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241. https://doi.org/10.1177/1745691612460685
Feeney, A. (2007). Individual differences, dual processes, and induction. In A. Feeney & E. Heit (Eds.), *Inductive reasoning: Experimental, developmental and computational approaches* (pp. 302–327). Cambridge University Press.
Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15. https://doi.org/10.1111/1467-8721.00158
Ferreira, F., & Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass*, *1*(1–2), 71–83. https://doi.org/10.1111/j.1749-818X.2007.00007.x
Fracasso, L. E., Bangs, K., & Binder, K. S. (2016). The contributions of phonological and morphological awareness to literacy skills in the adult basic education population. *Journal of Learning Disabilities*, *49*(2), 140–151. https://doi.org/10.1177/0022219414538513
Frankish, K. (2010). Dual-process and dual-system theories of reasoning. *Philosophy Compass*, *5*(10), 914–926. https://doi.org/10.1111/j.1747-9991.2010.00330.x
Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732
Freed, E. M., Hamilton, S. T., & Long, D. L. (2017). Comprehension in proficient readers: The nature of individual variation. *Journal of Memory and Language*, *97*, 135–153. https://doi.org/10.1016/j.jml.2017.07.008
Gagné, C. L., Spalding, T. L., & Nisbet, K. A. (2016). Processing English compounds: Investigating semantic transparency. *SKASE Journal of Theoretical Linguistics*, *13*(2).
Gagné, C. L., Spalding, T. L., & Schmidtke, D. (2019). LADEC: The large database of English compounds. *Behavior Research Methods*, *51*(5), 2152–2179. https://doi.org/10.3758/s13428-019-01282-6
Geva, E., & Ryan, E. B. (1985). Use of conjunctions in expository texts by skilled and less skilled readers. *Journal of Reading Behavior*, *17*(4), 331–346. https://doi.org/10.1080/10862968509547549
Goldman, S. R., & Murray, J. D. (1992). Knowledge of connectors as cohesion devices in text: A comparative study of native-english and english-as-a-second-language speakers. *Journal of Educational Psychology*, *84*(4), 504. https://doi.org/10.1037/0022-0663.84.4.504

Gu¨nther, F., & Marelli, M. (2019). Enter sandman: Compound processing and semantic transparency in a compositional perspective. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(10), 1872. https://doi.org/10.1037/xlm0000677

Guo, Y., Roehrig, A. D., & Williams, R. S. (2011). The relation of morphological awareness and syntactic awareness to adults' reading comprehension: Is vocabulary knowledge a mediating variable? *Journal of Literacy Research*, *43*(2), 159–183. https://doi.org/10.1177/1086296X11403086

Hobbs, J. R. (1979). Coherence and coreference. *Cognitive Science*, *3*(1), 67–90. https://doi.org/10.1207/s15516709cog0301_4

Holyoak, K. J. (2013). *Analogy and relational reasoning*. The Oxford Handbook of Thinking and Reasoning.

Juhasz, B. J. (2018). Experience with compound words influences their processing: An eye movement investigation with English compound words. *The Quarterly Journal of Experimental Psychology*, *71*(1), 103–112. https://doi.org/10.1080/17470218.2016.1253756

K¨ohne-Fuetterer, J., Drenhaus, H., Delogu, F., & Demberg, V. (2021). The online processing of causal and concessive discourse connectives. *Linguistics*, *59*(2), 417–448. https://doi.org/10.1515/ling-2021-0011

Kim, S. Y., Yap, M. J., & Goh, W. D. (2019). The role of semantic transparency in visual word recognition of compound words: A megastudy approach. *Behavior Research Methods*, *51*(6), 2722–2732. https://doi.org/10.3758/s13428-018-1143-3

Kleijn, S., Pander Maat, H. L., & Sanders, T. J. (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

Knoepke, J., Richter, T., Isberner, M.-B., Naumann, J., Neeb, Y., & Weinert, S. (2017). Processing of positive-causal and negative-causal coherence relations in primary school children and adults: A test of the cumulative cognitive complexity approach in German. *Journal of Child Language*, *44*(2), 297–328. https://doi.org/10.1017/S0305000915000872

Kotzer, M., Kirby, J. R., & Heggie, L. (2021). Morphological awareness predicts reading comprehension in adults. *Reading Psychology*, *42*(3), 302–322. https://doi.org/10.1080/02702711.2021.1888362

Levesque, K. C., Kieffer, M. J., & Deacon, S. H. (2017). Morphological awareness and reading comprehension: Examining mediating factors. *Journal of Experimental Child Psychology*, *160*, 1–20. https://doi.org/10.1016/j.jecp.2017.02.015

Libben, G., Gibson, M., Yoon, Y. B., & Sandra, D. (2003). Compound fracture: The role of semantic transparency and morphological headedness. *Brain and Language*, *84*(1), 50–64. https://doi.org/10.1016/S0093-934X(02)00520-5

Marelli, M., & Luzzatti, C. (2012). Frequency effects in the processing of Italian nominal compounds: Modulation of headedness and semantic transparency. *Journal of Memory and Language*, *66*(4), 644–664. https://doi.org/10.1016/j.jml.2012.01.003

Martin-Chang, S. L., & Gould, O. N. (2008). Revisiting print exposure: Exploring differential links to vocabulary, comprehension and reading rate. *Journal of Research in Reading*, *31*(3), 273–284. https://doi.org/10.1111/j.1467-9817.2008.00371.x

McCarron, S. P., & Kuperman, V. (2021). Is the author recognition test a useful metric for native and non-native english speakers? an item response theory analysis. *Behavior Research Methods*, *53*(5), 2226–2237. https://doi.org/10.3758/s13428-021-01556-y

Meyer, A., Zhou, E., & Frederick, S. (2018). The non-effects of repeated exposure to the cognitive reflection test. *Judgment and Decision Making*, *13*(3), 246–259. https://doi.org/10.1017/S1930297500007683

Millis, K. K., & Just, M. A. (1994). The influence of connectives on sentence comprehension. *Journal of Memory and Language*, *33*(1), 128–147. https://doi.org/10.1006/jmla.1994.1007

Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin*, *137*(2), 267. https://doi.org/10.1037/a0021890

Murray, J. D. (1997). Connectives and narrative text: The role of continuity. *Memory & Cognition*, *25*(2), 227–236. https://doi.org/10.3758/BF03201114

Nation, K., & Snowling, M. J. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading*, *27*(4), 342–356. https://doi.org/10.1111/j.1467-9817.2004.00238.x

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

Nippold, M. A., Schwarz, I. E., & Undlin, R. A. (1992). Use and understanding of adverbial conjuncts: A developmental study of adolescents and young adults. *Journal of Speech, Language, and Hearing Research*, *35*(1), 108–118. https://doi.org/10.1044/jshr.3501.108

Otero, I., Salgado, J. F., & Moscoso, S. (2022). Cognitive reflection, cognitive intelligence, and cognitive abilities: A meta-analysis. *Intelligence*, *90*, 101614. https://doi.org/10.1016/j.intell.2021.101614

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227–247). Blackwell Publishing. https://doi.org/10.1002/9780470757642.ch13

Pollatsek, A., & Hy¨on¨a, J. (2005). The role of semantic transparency in the processing of Finnish compound words. *Language and Cognitive Processes*, *20*(1–2), 261–290. https://doi.org/10.1080/01690960444000098

Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A., & Hamilton, J. (2016). The development and testing of a new version of the cognitive reflection test applying item response theory (IRT). *Journal of Behavioral Decision Making*, *29*(5), 453–469. https://doi.org/10.1002/bdm.1883

R Development Core Team. (2008). R: A language and environment for statistical computing [Computer software manual]. http://www.R-project.org

Raven, J. C., & Court, J. (1938). *Raven's progressive matrices*. Western Psychological Services.

Sanders, T. J. M., & Noordman, L. G. M. (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse Processes*, *29*(1), 37–60. https://doi.org/10.1207/S15326950dp2901_3

Sanders, T. J. M., Spooren, W. P. M. S., & Noordman, L. G. M. (1992). Toward a taxonomy of coherence relations. *Discourse Processes*, *15*(1), 1–35. https://doi.org/10.1080/01638539209544800

Sandra, D. (1990). On the representation and processing of compound words: Automatic access to constituent morphemes does not occur. *The Quarterly Journal of Experimental Psychology Section A*, *42*(3), 529–567. https://doi.org/10.1080/14640749008401236

Scholman, M. C. J., Demberg, V., & Sanders, T. J. (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

Scholz, M., & Scheer, D. (2020). The relationship between reading skills and intelligence in students with and without special educational needs in learning. *Frontiers in Education*, *5*, 123. https://doi.org/10.3389/feduc.2020.00123

Sirota, M., & Juanchich, M. (2018). Effect of response format on cognitive reflection: Validating a twoand four-option multiple choice question version of the cognitive reflection test. *Behavior Research Methods*, *50*(6), 2511–2522. https://doi.org/10.3758/s13428-018-1029-4

Smolka, E., & Libben, G. (2017). 'Can you wash off the hogwash?'–semantic transparency of first and second constituents in the processing of German compounds. *Language, Cognition and Neuroscience*, *32*(4), 514–531. https://doi.org/10.1080/23273798.2016.1256492

Spencer, M., Muse, A., Wagner, R. K., Foorman, B., Petscher, Y., Schatschneider, C., Tighe, E. L., & Bishop, M. D. (2015). Examining the underlying dimensions of morphological awareness and vocabulary knowledge. *Reading and Writing*, *28*(7), 959–988. https://doi.org/10.1007/s11145-015-9557-0

Stanovich, K. E. (1999). *Who is rational?: Studies of individual differences in reasoning*. Psychology Press.

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, *24*(4), 402–433. https://doi.org/10.2307/747605

Stanovich, K. E., & West, R. F. (2008). On the relative Independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*, *94*(4), 672. https://doi.org/10.1037/0022–3514.94.4.672

Stieger, S., & Reips, U.-D. (2016). A limitation of the cognitive reflection test: Familiarity. *PeerJ*, *4*, e2395. https://doi.org/10.7717/peerj.2395

Swanson, H. L., Howard, C. B., & Saez, L. (2006). Do different components of working memory underlie different subgroups of reading disabilities? *Journal of Learning Disabilities*, *39*(3), 252–269. https://doi.org/10.1177/00222194060390030501

Thompson, V. A. (2009). *Dual-process theories: A metacognitive perspective*. Oxford University Press.

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making*, *11*(1), 99. https://doi.org/10.1017/S1930297500007622

Tighe, E. L., & Binder, K. S. (2015). An investigation of morphological awareness and processing in adults with low literacy. *Applied Psycholinguistics*, *36*(2), 245–273. https://doi.org/10.1017/S0142716413000222

Tiu, R. D., Jr, Thompson, L. A., & Lewis, B. A. (2003). The role of IQ in a component model of reading. *Journal of Learning Disabilities*, *36*(5), 424–436. https://doi.org/10.1177/00222194030360050401

Tong, X., Deacon, S. H., Kirby, J. R., Cain, K., & Parrila, R. (2011). Morphological awareness: A key to understanding poor reading comprehension in English. *Journal of Educational Psychology*, *103*(3), 523. https://doi.org/10.1037/a0023495

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the cognitive reflection test. *Thinking & Reasoning*, *20*(2), 147–168. https://doi.org/10.1080/13546783.2013.844729

Tran, A. H., Tremblay, K. A., & Binder, K. S. (2020). The factor structure of vocabulary: An investigation of breadth and depth of adults with low literacy skills. *Journal of Psycholinguistic Research*, *49*(2), 335–350. https://doi.org/10.1007/s10936-020-09694-8

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. (2022). 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

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

Vermeiren, H., Vandendaele, A., & Brysbaert, M. (2023). Validated tests for language research with university students whose native language is English: Tests of vocabulary, general knowledge, author recognition, and reading comprehension. *Behavior Research Methods*, *55*, 1036–1068. https://doi.org/10.3758/s13428-022-01856-x

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

Webber, B., Prasad, R., Lee, A., & Joshi, A. (2019). *The Penn Discourse Treebank 3.0 annotation manual*. University of Pennsylvania.

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*(S1), S42–S65. https://doi.org/10.1111/1467-9817.12090

Welsh, M. B. (2022). What is the CRT? intelligence, personality, decision style or attention? In *Proceedings of the annual meeting of the cognitive science society* (Vol. 44).

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

Woike, J. K. (2019). Upon repeated reflection: Consequences of frequent exposure to the Cognitive Reflection Test for Mechanical Turk participants. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02646

Xiang, M., & Kuperberg, G. (2015). Reversing expectations during discourse comprehension. *Language, Cognition and Neuroscience*, *30*(6), 648–672. https://doi.org/10.1080/23273798.2014.995679

Xu, X., Chen, Q., Panther, K.-U., & Wu, Y. (2018). Influence of concessive and causal conjunctions on pragmatic processing: Online measures from eye movements and self-paced reading. *Discourse Processes*, *55*(4), 387–409. https://doi.org/10.1080/0163853X.2016.1272088

Zufferey, S., & Gygax, P. (2020a). Do teenagers know how to use connectives from the written mode? *Lingua*, *234*, 102779. https://doi.org/10.1016/j.lingua.2019.102779

Zufferey, S., & Gygax, P. (2020b). "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

Zwitserlood, P. (1994). The role of semantic transparency in the processing and representation of dutch compounds. *Language and Cognitive Processes*, *9*(3), 341–368. https://doi.org/10.1080/01690969408402123

Appendix

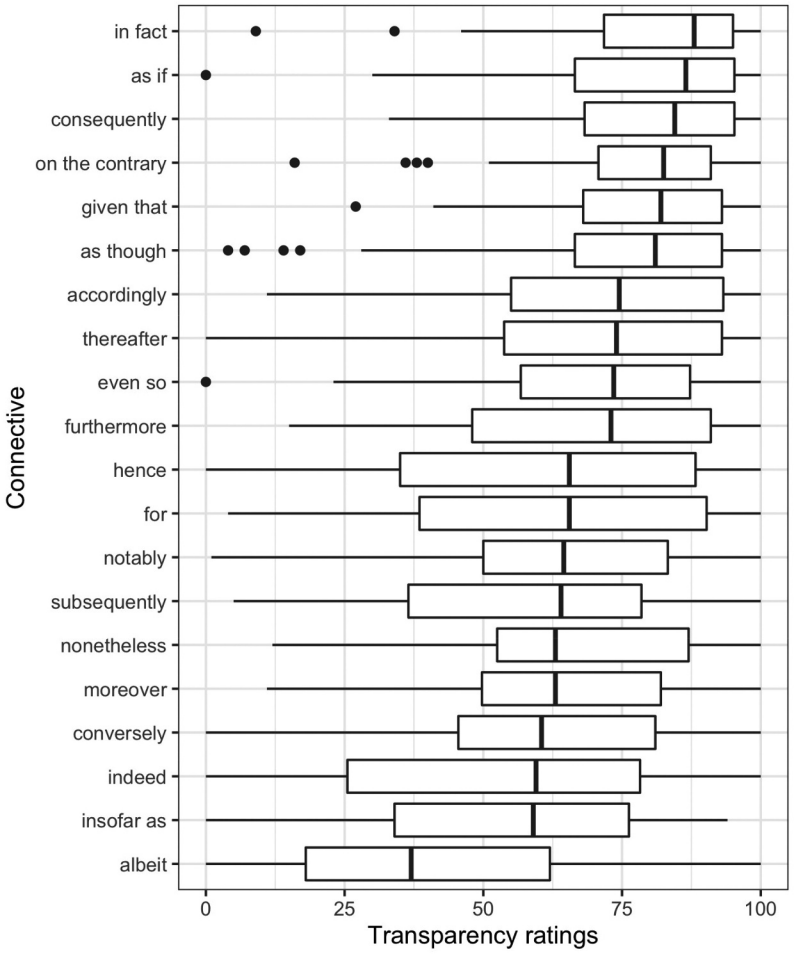

Figure 1A. Presents a distribution of the lexical transparency ratings per connective.