Searching for signals: Readers' sensitivity to signals for discourse relations

Dissertation
zur Erlangung des akademischen Grades
eines Doktors der Philosophie
der Philosophischen Fakultät
der Universität des Saarlandes

vorgelegt von
Anna Maria Marchal
aus Zeist

Saarbrücken, 07. April 2024

Dekan der Fakultät P: Prof. Dr. Nine Miedema
Erstberichterstatter: Vera Demberg
Zweitberichterstatter: Matthew Crocker
Drittberichterstatter: Jet Hoek
Tag der letzten Prüfungsleistung: 13. März 2025

Abstract

For comprehension to be successful, readers and listeners need to understand the meaning of individual words and sentences, but also have to know how these words and sentences are related to each other. That is, comprehenders need to establish a *coherent* mental representation of the discourse (Sanders et al., 1992; Zwaan & Rapp, 2006; Van den Broek, 2010). Discourse relations, which refer to the relations between segments in a text (Hobbs, 1979; Sanders et al., 1992; Zufferey & Degand, 2024), are an important part of such a mental representation. Readers and listeners can infer these relations based on linguistic information (e.g. connectives) as well as extra-linguistic information (e.g. world knowledge). This dissertation set out to investigate to what extent readers use these different types of information. Specifically, we examined four factors that could influence how readers exploit linguistic signals for discourse relations: characteristics of the linguistic signal, of the discourse relation, of the reader and of the language.

Connectives, such as *because* or *but*, are the most salient linguistic signals for discourse relations and have been shown to help readers to process the discourse relation (e.g. Cozijn et al., 2011; Kleijn et al., 2019; Köhne-Fuetterer et al., 2021). However, most relations are signaled by linguistic cues other than connectives (cf. Das & Taboada, 2018b) and much less is known about whether readers exploit these signals. We discuss five features in which connective and non-connective cues differ and argue that readers' sensitivity to linguistic signals depends on the salience and informativeness of the cue. Furthermore, we extend previous research on the role of linguistic cues in discourse relation processing, by investigating a non-lexical signal of discourse relations, showing that such a cue can influence readers' off-line expectations about upcoming discourse relations.

Secondly, to what extent readers rely on linguistic information may depend on the discourse relation. Discourse relations have been shown differ with respect to their processing difficulty (e.g. Sanders & Noordman, 2000). In line with the *causality-by-default* hypothesis (Sanders, 2005), we find evidence that the presence of a connective facilitates reading more in non-causal than in causal relations. In addition, we show that the processing difficulty of a relation is dependent on how predictable the relation type and its content is. However, we do not find evidence that predictability of the relation influences whether readers rely on the presence of a connective.

Thirdly, we provide evidence that readers draw on their domain knowledge when inferring discourse relations. The availability of domain knowledge was also found to

influence whether readers can exploit non-connective signals for discourse relations, since these signals sometimes require domain knowledge. Finally, we explored whether the use of non-linguistic signals depends on language typology. We hypothesized that speakers of synthetic languages would rely more on the presence of linguistic signals for discourse relations than speakers of analytic languages, but find no evidence for this.

In sum, readers draw on both linguistic and non-linguistic information to establish a coherent mental representation. In addition, the research in this dissertation shows that the extent to which readers exploit linguistic cues depends on characteristics of the signal, the discourse relation and the reader. By investigating the processes involved in establishing coherence, this research provides theoretical insights into language understanding and human cognition, but can also inform research on how to improve readers' text comprehension as well as the readability of texts.

Summary

Language is a central aspect of everyday life. We use language to tell jokes to our loved ones, listen to a friend's ideas, read a news item about recent elections and reply to an email from a coworker. These processes often seem effortless, but producing and understanding language is a complex process, which is an ability that is unique to humans. To illustrate, to comprehend this written sentence, one does not only need to decode the sequences of black symbols (i.e. letters) into meaningful words, but also needs to understand how these words are structured to form a meaningful sentence. In addition, in longer excerpts of a text, also referred to as a *discourse*, comprehenders have to understand how the different sentences are related to each other and how the concepts described in the text are related to existing knowledge. Only then does the discourse make sense.

Theories of discourse comprehension assume that readers and listeners construct a mental representation of the text (Zwaan & Rapp, 2006; Van den Broek, 2010). For comprehension to be successful, this mental representation needs to be *coherent* (Hobbs, 1979; Sanders et al., 1992; Kehler, 2006). That is, the parts of the mental representation should be connected in a meaningful way. Even when the parts of a text are seemingly unrelated, readers still try to establish coherence (Hobbs, 1979). Consider the following discourse: *Mary is cycling to work. She loves pizza.* At first sight, this does not make sense, but comprehenders will still try to understand this text. They might assume that Mary's love of pizza has resulted in her eating a lot lately and that she now tries to compensate for these calories by cycling to work. They have now established coherence by inferring how these sentences are related: Mary's love for pizza provides the *reason* for why she is cycling to work. These relations between sentences in a discourse, or their mental representations, are referred to as *discourse relations* (Sanders et al., 1992; Zufferey & Degand, 2024) and the parts of the discourse relation are called *arguments*. Discourse relations are the focus of the research in this dissertation.

How do readers establish coherence? One way in which readers do this is by using linguistic signals. The most salient signals for discourse relations are connectives, such as *because*, *but*, and *therefore*. They have been researched extensively, with prior work showing that they help readers to process the discourse relation (e.g. Cozijn et al., 2011), predict upcoming material (e.g. Köhne-Fuetterer et al., 2021) and understand the text better (e.g. Kleijn et al., 2019). However, there are also other cues that can help readers to infer what relation holds between sentences. To illustrate, in the

discourse *Mary loves pizza. John hates it.*, the verbs are antonyms, signaling that the two sentences are in contrast to each other. Compared to connectives, much less is known about the role of these non-connective signals in discourse processing and comprehension. Furthermore, readers could rely on their background knowledge to understand the relation. For example, in the discourse *Mary got very tanned. She went on vacation to Greece.*, the causal relation between tanning and going to the Greece can be derived based on world knowledge that Greece is generally very sunny.

Comprehenders can thus use linguistic and extra-linguistic sources of information to infer how the parts of a discourse are related. However, it is still unclear whether readers always do so. For example, do speakers of different languages rely similarly on connectives as cues for the discourse relation? And do readers also exploit non-connective signals when establishing coherence? This dissertation set out to investigate **which factors influence readers' sensitivity to linguistic signals of discourse relations**. More specifically, we explored four factors that could influence how readers use linguistic information when inferring discourse relations: characteristics of the signal, of the discourse relation, of the reader and of the language. We present a review of earlier work as well as four empirical studies to investigate these factors.

Chapter 3 reviews previous literature on the nature of linguistic signals and their role in the representation and processing to investigate how **characteristics of the signal** influence readers' sensitivity to a discourse cue. We define a discourse relation signal (or a discourse cue) as any linguistic element that provides information about the discourse relation. We show that these signals differ with respect to various features. Firstly, some discourse cues, such as connectives, are specialized for signaling discourse relations and do not contribute to the truth-conditions of the arguments. Others, such as antonyms, do not only provide information about the discourse relation, but also have propositional meaning. Secondly, some cues are more informative about which discourse relation is signaled than others. For example, *because* is only used in `result` relations, whereas *while* can be used in both `synchronous` as well as `contrast` relations. Third, discourse cues differ in whether their form is context-dependent. Connectives are grammaticalized and are thus immutable, but this is not the case for cue phrases like *for this reason*, *due to the weather* or antonyms. Fourth, the examples of discourse cues so far are all lexical. However, this does not need to be the case. Syntactic structure (Crible & Pickering, 2020) and prosody (Hu et al., 2023) have also been argued to signal differences between relation types. Finally, linguistic cues that are not specialized in signaling the discourse relation have a different

primary meaning. This meaning may be similar to that of the discourse relation, such as negation in the case of `contrast` relations, but it may also be unrelated, as in the case of certain syntactic structures. Based on findings from previous literature, we argue that these features influence how sensitive comprehenders are to the linguistic signal. The more salient and informative the discourse cue, the stronger its effect on discourse processing and representation.

This hypothesis is supported by the findings from two empirical studies presented in this dissertation. In Chapter 6, we examine readers' sensitivity to a non-specialized, non-lexical cue for discourse relations: gerund free adjuncts. Gerund free adjuncts are subordinate clauses that start with a present participle, as in *Painting his house, Mo wore an old sweater.* A corpus investigation showed that gerund free adjuncts often occur in `result` relations. In addition, readers expected more `result` relations when provided with a gerund free adjunct in a continuation task. However, we did not find evidence that readers are sensitive to this type of cue in other tasks. In a paraphrase selection task, readers' preference for gerund free adjuncts was not found to depend on the type of relation (i.e. `result` or `specification`). In addition, a self-paced reading experiment did not show facilitation of the presence of a gerund free adjunct when reading `result` relations. This finding contrasts with those from Chapter 5, which shows that the presence of a connective does lead to faster processing of `result` relations. However, compared to gerund free adjuncts, connectives are more informative and more specialized cues that are also immutable and lexical.

In Chapter 4, we investigated how readers' sensitivity to connectives depends on **characteristics of the discourse relation**. More specifically, we conducted two self-paced reading experiments comparing the facilitating effect of the presence of a connective in `result` relations with `contrast` relations (Experiment 1) and with `concession` relations (Experiment 2). Contrary to our expectations based on earlier work, we found no effect of the presence of a connective in either relation in Experiment 1, possibly due to methodological limitations. In Experiment 2, however, the relation was read faster when a connective was present compared to when it was absent. This effect interacted with relation type. More specifically, the connective facilitated reading in `concession` relations, but not in `result` relations. This suggests that readers rely more on the connective when the relation is more difficult to infer without it. For `result` relations, readers might arrive at this interpretation regardless of whether there is a connective (cf. causality-by-default hypothesis Sanders, 2005).

The experiments in Chapter 4 also investigated whether the effect of the connective on reading was dependent on **characteristics of the language**, by comparing native

speakers of English and German. Synthetic languages, such as German, have more inflectional morphology and a higher morpheme-to-word ratio than analytic languages, such as English. To illustrate, German, but not English, encodes differences between first and second person in verbs and case in nouns. In synthetic languages, meaning is more often encoded in the linguistic signal, whereas speakers of analytic languages more often have to infer the meaning from the context. We therefore hypothesized that speakers of German would be more sensitive to the presence of a connective than speakers of English (cf. Blumenthal-Dramé, 2021). However, we found no evidence for this hypothesis in either of the two experiments.

In Chapter 5, we investigated whether readers' sensitivity to the presence of a connective depends on another **characteristic of the relation**: the predictability of the relation. Previous research has shown that comprehenders continuously make predictions during language processing (e.g. Altmann & Kamide, 1999; Heilbron et al., 2022) and that the processing difficulty of a word is proportional to its unexpectedness (cf. Levy, 2008; Wilcox et al., 2023). We investigated whether this also applies to the processing of discourse relations. More specifically, we examined two types of unexpectedness: that of the relation type (i.e. whether the relation was a `result` or not) and that of the content of the relation. To illustrate, when hearing *Angela hadn't paid rent for months*, readers might have expectations about whether the speaker is next going to talk about why Angela hadn't paid rent (i.e. a `reason` relation) or about what the consequences of these payment arrears are (i.e. a `result` relation). In other words, they predict the relation type. In addition, comprehenders might predict *what* such consequences might be (e.g. a visit from an angry landlord or being evicted). This is a prediction about the content. We hypothesized that the difficulty of processing *Angela was evicted* would be proportional to the unexpectedness of the relation type and of the content given the context. In addition, we aimed to investigate whether this would explain the facilitating effect of the connective: the connective might reduce the unexpectedness of the upcoming relation type and content and as such reduce processing difficulty.

In a continuation pretest, we indeed show that `result` relation types are more expected when a connective is present and that this helps readers to make more accurate predictions about the content. In a subsequent self-paced reading and eye-tracking-while-reading experiment, we found evidence that more predictable content is read faster. Also, when the `result` relation was more expected, the relation was also read faster, but only when this allowed for more accurate predictions about the content. Surprisingly, we found that first-pass reading times were longer for

more predictable `result` relations, when controlling for other types of unexpectedness (e.g. about the content of the relation). Possibly, readers confirm their relation type prediction when their content prediction is not borne out. In addition, the facilitating effect of the connective was independent from effects of unexpectedness. In other words, we did not find evidence that readers' sensitivity to the presence of a connective depends on how predictable the relation type or content is.

How **characteristics of the reader** influence whether readers exploit linguistic signals was examined in Chapter 7. More specifically, the goal of this chapter was to investigate whether domain knowledge affects readers' interpretation of discourse relations and their reliance on linguistic signals. Experts in the fields of biomedical sciences or economics were asked to insert connectives in texts from their own domain (e.g. biomedical experts reading biomedical research papers) as well as the other domain (e.g. biomedical experts reading financial newspapers). These connectives were categorized into relation classes to examine readers' accuracy in inferring the target relation. Biomedical experts were more accurate than economics experts on inferring relations in the biomedical texts, showing that readers use their domain knowledge to infer discourse relations. No such difference was found for the interpretation of relations in the financial newspapers, probably because these are aimed at a broader audience than biomedical research papers. When the relation cannot be inferred based on existing knowledge, readers make underspecified interpretations. Furthermore, we show that readers exploit linguistic signals for discourse relations, but that such cues sometimes require domain knowledge. For example, to understand that two concepts are antonyms, readers first need to know what those concepts are. However, even when these cues only require general knowledge, low-knowledge readers did not always take advantage of them. This suggests that non-connective cues might only be used to confirm discourse relation interpretation.

The research in this dissertation provides valuable insights into various theories of language processing. First of all, with respect to discourse processing, the findings reveal that readers exploit both linguistic (connectives and non-connective cues) and extra-linguistic (background knowledge) sources of information to establish coherence. We also show that readers' sensitivity to signals of discourse relations depends on characteristics of the signal, the relation and the reader. More specifically, the effect of a linguistic cue is argued to be stronger when the signal is more salient and more informative, the relation is non-causal and the reader can rely on existing knowledge to confirm the signal's meaning. Second, the research presented here contributes to information-theoretic accounts of language processing by showing that

the unexpectedness of semantic content and discourse structure influence processing difficulty, but do not explain the facilitating effect of the connective. Third, we contribute to research on statistical learning by revealing that readers are aware of correlations between syntactic structure and discourse-level meaning. Fourth, with respect to research on individual differences, the work in this dissertation shows that the interpretation of discourse relations, and the use of linguistic signals, depends on the readers' background knowledge. Fifth, we extend cross-linguistic research on language processing and find that there are cross-linguistic similarities in the use of connectives for discourse relation processing. Finally, this dissertation highlights the importance of converging evidence, by showing that combining evidence from various methodologies can facilitate theory building.

Zusammenfassung

Sprache ist ein zentraler Aspekt des täglichen Lebens. Wir verwenden Sprache, um unseren Freunden Witze zu erzählen, den Ideen eines Freundes zuzuhören, einen Artikel über die letzten Wahlen zu lesen und eine E-Mail von einem Kollegen zu beantworten. Diese Vorgänge scheinen uns oft mühelos, aber Sprache zu produzieren und zu verstehen ist ein komplizierter Prozess, und eine Fähigkeit, die nur dem Menschen eigen ist. Um zum Beispiel einen geschriebenen Satz zu verstehen, muss man nicht nur eine Abfolge an Buchstaben in sinnvolle Wörter entschlüsseln, sondern auch verstehen, wie diese Wörter strukturiert sind, um einen sinnvollen Satz zu bilden. Bei längeren Textausschnitten, die auch als *Diskurs* bezeichnet werden, müssen die Leser außerdem verstehen, wie die verschiedenen Sätze miteinander in Relation stehen und wie die im Text beschriebenen Konzepte mit dem vorhandenen Weltwissen zusammenhängen. Erst dann ergibt der Diskurs einen Sinn.

Theorien des Diskursverständnisses gehen davon aus, dass Leser und Hörer eine mentale Repräsentation des Textes aufbauen. Damit das Verstehen erfolgreich ist, muss diese mentale Repräsentation *kohärent* sein (Hobbs, 1979; Sanders et al., 1992; Kehler, 2006). Dies bedeutet dass die Teilaspekte der mentalen Repräsentation auf sinnvolle Weise miteinander verbunden sein sollten. Selbst wenn die Teilaspekte eines Textes scheinbar nicht miteinander verbunden sind, versuchen die Leser dennoch, Kohärenz herzustellen (Hobbs, 1979). Betrachten Sie den folgenden Diskurs: *Anna fährt mit dem Fahrrad zur Arbeit. Sie liebt Pizza.* Auf den ersten Blick ergibt das keinen Sinn, aber der Leser wird trotzdem versuchen, diesen Text zu verstehen. Er könnte annehmen, dass Annas Vorliebe für Pizza dazu geführt hat, dass sie in letzter Zeit sehr viel gegessen hat und nun versucht, diese Kalorien auszugleichen, indem sie mit dem Fahrrad zur Arbeit fährt. So hat er Kohärenz hergestellt, indem er hergeleitet hat, wie diese Sätze zusammenhängen: Annas Liebe zu Pizza ist der Grund, warum sie mit dem Fahrrad zur Arbeit fährt. Diese Relationen zwischen Sätzen in einem Diskurs bzw. ihre mentalen Repräsentationen werden als *Diskursrelationen* bezeichnet (Sanders et al., 1992; Zufferey & Degand, 2024) und die Teile der Diskursrelation werden *Argumente* genannt. Diskursrelationen sind der Schwerpunkt der Forschung in dieser Dissertation.

Wie stellen die Leser Kohärenz her? Eine Möglichkeit, wie Leser dies tun, ist die Verwendung sprachlicher Signale. Die auffälligsten Signale für Diskursrelationen sind Konnektive wie *weil*, *aber* und *deshalb*. Sie wurden ausgiebig erforscht und frühere Arbeiten haben gezeigt, dass sie Lesern helfen, die Diskursrelation zu ve-

rarbeiten (z.B., Cozijn et al., 2011), bevorstehendes Material vorherzusagen (z.B., Köhne-Fuetterer et al., 2021) und einen Text besser zu verstehen (z.B., Kleijn et al., 2019). Es gibt jedoch auch andere Anhaltspunkte, die dem Leser helfen können, eine Relation zwischen den Sätzen zu erkennen. Zur Veranschaulichung: Im Diskurs *Anna liebt Pizza. John hasst sie.* sind die Verben Antonyme, die signalisieren, dass die beiden Sätze im Gegensatz zueinander stehen. Im Vergleich zu Konnektiven ist über die Rolle dieser nicht-konnektiven Signale bei der Verarbeitung und dem Verständnis von Diskurs weit weniger bekannt. Schließlich könnten die Leser auf ihr Hintergrundwissen zurückgreifen, um die Relation zu verstehen. Zum Beispiel in dem Diskurs *Anna wurde sehr braun. Sie fuhr in den Urlaub nach Griechenland.* kann die kausale Relation zwischen Bräunung und Urlaub in Griechenland aus dem Wissen abgeleitet werden, dass Griechenland im Allgemeinen sehr sonnig ist.

Leser können also sprachliche und außersprachliche Informationsquellen nutzen, um daraus zu schließen, wie die Teile eines Diskurses zusammenhängen. Es ist jedoch noch unklar, ob Menschen dies immer tun. Verlassen sich zum Beispiel Sprecher verschiedener Sprachen in ähnlicher Weise auf Konnektive als Hinweise auf die Diskursrelation? Und nutzen Leser auch nicht-konnektive Signale, um Kohärenz herzustellen? In dieser Dissertation wurde daher untersucht, **welche Faktoren die Sensibilität der Leser für sprachliche Signale von Diskursrelationen beeinflussen**. Genauer gesagt wurden vier Faktoren untersucht, die Einfluss darauf haben könnten, wie Leser sprachliche Informationen über Diskursrelationen nutzen: Merkmale des Signals, der Diskursrelation, des Lesers und der Sprache. Zur Untersuchung dieser Faktoren präsentieren wir einen Überblick über frühere Arbeiten sowie vier empirische Studien.

Kapitel 3 gibt einen Überblick über die bisherige Literatur zur Natur sprachlicher Signale und ihrer Rolle bei der mentalen Repräsentation und Verarbeitung, um zu untersuchen, wie die **Eigenschaften des Signals** die Sensibilität der Leser für das Signal beeinflussen. Wir definieren Diskursrelationssignal als jedwedes sprachliche Element, das Informationen über die Diskursrelation liefert. Wir zeigen, dass sich diese Signale in Bezug auf verschiedene Merkmale unterscheiden. So sind einige Diskurssignale, wie z.B. Konnektive, auf die Signalisierung von Diskursrelationen spezialisiert und tragen nicht zu den Wahrheitsbedingungen der Argumente bei. Andere Diskurssignale, wie Antonyme, liefern nicht nur Informationen über die Diskursrelation, sondern haben auch propositionale Bedeutung. Zweitens sind einige Signale informativer darüber, welche Diskursrelation signalisiert wird, als andere. Zum Beispiel wird *weil* nur in kausalen `result`-Relationen verwendet, während *gleichzeitig* sowohl

in zeitlichen `synchronous`- als auch in negativen `contrast`-Relationen verwendet werden kann. Drittens unterscheiden sich Diskurssignale darin, ob ihre Form kontextabhängig ist. Konnektive werden grammatikalisiert und sind daher unveränderlich, wohingegen ganze Phrasen wie *aus diesem Grund*, *wegen des Wetters* oder Antonyme durchaus veränderlich sind. Viertens: Die bisherigen Beispiele für Diskurssignale sind alle lexikalisch. Dies muss jedoch nicht der Fall sein. Es ist bekannt, dass auch die syntaktische Struktur (Crible & Pickering, 2020) und die Prosodie (Hu et al., 2023) Unterschiede zwischen den Relationstypen signalisieren. Schließlich haben sprachliche Signale, die nicht auf die Signalisierung der Diskursrelation spezialisiert sind, eine andere primäre Bedeutung. Diese Bedeutung kann der der Diskursrelation ähnlich sein, wie z.B. die Negation im Fall von `contrast`-Relationen, sie kann aber auch unabhängig davon sein (wie im Fall bestimmter syntaktischer Strukturen). Auf der Grundlage von Erkenntnissen aus der bisherigen Literatur argumentieren wir, dass diese Eigenschaften beeinflussen, wie empfindlich Leser auf das sprachliche Signal reagieren. Je auffälliger und informativer das Diskurssignal ist, desto stärker ist seine Wirkung auf die Diskursverarbeitung und -repräsentation.

Diese Hypothese wird durch die Ergebnisse von zwei empirischen Studien gestützt. In Kapitel 6 untersuchten wir die Sensibilität der Leser für einen nicht-spezialisierten, nicht-lexikalischen Hinweis auf Diskursrelationen: freie Gerundien. Freie Gerundien sind Nebensätze, die mit einem Partizip Präsens beginnen, wie in *Painting his house, Mo wore an old sweater.* (*Mo trug beim Streichen seines Hauses einen alten Pullover.*) Eine Korpusuntersuchung zeigt, dass freie Gerundien häufig in `result`-Relationen vorkommen. Darüber hinaus erwarteten die Leser mehr `result`-Relationen, wenn sie in einer Fortsetzungsaufgabe ein freies Gerundium vorfanden. Wir fanden jedoch keine Hinweise darauf, dass die Leser bei anderen Aufgaben auf diese Art von Signalen reagieren. Bei einer Aufgabe zur Auswahl von Paraphrasen hing die Präferenz der Leser für freie Gerundien nicht von der Art der Relation (d. h. `result` oder `specification`) ab. Darüber hinaus zeigte ein self-paced reading Experiment nicht, dass das Vorhandensein eines freien Gerundiums beim Lesen von `result`-Relationen förderlich ist. Diese Ergebnisse stehen im Gegensatz zu denen aus Kapitel 5, die zeigen, dass das Vorhandensein eines Konnektivs zu einer schnelleren Verarbeitung von `result`-Relationen führt. Im Gegensatz zu freien Gerundien sind Konnektive jedoch informativere und speziellere Signale, die zudem unveränderlich und lexikalisch sind.

In Kapitel 4 haben wir untersucht, wie die Sensibilität der Leser für Konnektive von den **Eigenschaften der Diskursrelation** abhängt. Genauer gesagt haben wir zwei

self-paced reading Experimente durchgeführt, in denen wir den erleichternden Effekt des Vorhandenseins eines Konnektivs in `result`-Relationen mit `contrast`-Relationen (Experiment 1) und mit `concession`-Relationen (Experiment 2) verglichen haben. Entgegen unseren Erwartungen, die wir aus früheren Arbeiten ableiten konnten, fanden wir in Experiment 1 keinen Effekt des Konnektivs für eine der beiden Relationen, was möglicherweise auf methodische Limitationen zurückzuführen ist. In Experiment 2 wurde die Relation jedoch schneller gelesen, wenn ein Konnektiv vorhanden war, im Vergleich dazu wenn es nicht vorhanden war. Dieser Effekt hing vom Relationstyp ab. Genauer gesagt erleichterte das Konnektiv das Lesen bei `concession`-Relationen, nicht aber bei `result`-Relationen. Dies deutet darauf hin, dass sich die Leser mehr auf ein Konnektiv verlassen, wenn die Relation ohne das Konnektiv schwieriger zu erschließen ist. Bei `result`-Relationen könnten die Leser zu dieser Interpretation gelangen, unabhängig davon, ob ein Konnektiv (vgl. causality-by-default hypothesis Sanders, 2005) vorhanden ist.

In den Experimenten in Kapitel 4 wurde auch untersucht, ob die Wirkung eines Konnektivs unterschiedlich auf das Lesen im Deutschen im Vergleich zum Englischen ist (d. h. **Eigenschaften der Sprache**). Synthetische Sprachen, wie das Deutsche, haben mehr Flexionsmorphologie und ein höheres Morphem-Wort-Verhältnis als analytische Sprachen, wie das Englische. So kodiert das Deutsche, nicht aber das Englische, Unterschiede zwischen der ersten und zweiten Person bei Verben und dem Kasus bei Substantiven. In synthetischen Sprachen ist die Bedeutung häufiger im sprachlichen Signal kodiert, während Sprecher analytischer Sprachen die Bedeutung häufiger aus dem Kontext erschließen müssen. Wir stellten daher die Hypothese auf, dass Sprecher des Deutschen stärker auf das Vorhandensein eines Konnektivs reagieren würden als Sprecher des Englischen (Blumenthal-Dramé, 2021). Wir fanden jedoch in keinem der beiden Experimente Belege für diese Hypothese.

In Kapitel 5 haben wir untersucht, ob die Sensibilität der Leser für das Vorhandensein eines Konnektivs von einer anderen **Eigenschaft der Relation**, der Vorhersagbarkeit der Relation, abhängt. Frühere Studien haben gezeigt, während des Verstehens von Sprache ständig Vorhersagen getroffen werden (z.B. Altmann & Kamide, 1999; Heilbron et al., 2022) und dass die Verarbeitungsschwierigkeit eines Wortes proportional zu seiner Unerwartetheit ist (vgl. Levy, 2008; Wilcox et al., 2023). Wir untersuchten, ob dies auch für die Verarbeitung von Diskursrelationen gilt. Genauer gesagt untersuchten wir zwei Arten von Unerwartetheit: die des Relationstyps (d. h., ob die Relation ein `result` war oder nicht) und die des Inhalts der Relation. Zur Veranschaulichung: Wenn man *Angela hatte monatelang die Miete nicht*

bezahlt liest, könnten die Leser Erwartungen darüber haben, ob der Sprecher als nächstes darüber sprechen wird, warum Angela die Miete nicht bezahlt hat (d. h. eine `reason`-Relation) oder darüber, was die Konsequenzen dieses Zahlungsrückstands sind (d. h. eine `result`-Relation). Mit anderen Worten, sie sagen den Relationstyp voraus. Darüber hinaus könnten Leser vorhersagen, was solche Folgen sein könnten (z.B. der Besuch eines verärgerten Vermieters oder eine Zwangsräumung). Dabei handelt es sich um eine Vorhersage über den Inhalt der Relation. Wir stellten die Hypothese auf, dass die Schwierigkeit der Verarbeitung von *Angela wurde zwangsgeräumt* proportional zur Unerwartetheit des Relationstyps und des Inhalts angesichts des Kontexts sein würde. Darüber hinaus wollten wir untersuchen, ob dies den erleichternden Effekt des Konnektivs erklären würde: Das Konnektiv könnte die Unerwartetheit des bevorstehenden Relationstyps und -inhalts verringern und damit die Verarbeitungsschwierigkeiten reduzieren. In einem Experiment zur Satzvervollständigung konnten wir tatsächlich zeigen, dass `result`-Relationen eher erwartet werden, wenn ein Konnektiv vorhanden ist, und dass dies den Lesern hilft, genauere Vorhersagen über den Inhalt zu treffen. In einem anschließenden self-paced reading Experiment und Eye-Tracking fanden wir Hinweise darauf, dass vorhersehbare Inhalte schneller gelesen werden. Wenn die Relation `result` erwarteter war, wurde die Relation ebenfalls schneller gelesen, aber nur, wenn diese genauere Vorhersagen über den Inhalt ermöglichte. Überraschenderweise stellten wir fest, dass *first-pass* Lesezeiten bei vorhersehbareren `result`-Relationen länger waren, wenn man andere Arten von Unerwartetheit (z.B. über die Inhalt des Relations) berücksichtigt. Möglicherweise wollen die Leser ihre Vorhersage über die Art der Relation bestätigen, wenn sich ihre Vorhersage über den Inhalt nicht bestätigte. Darüber hinaus war der erleichternde Effekt des Konnektivs unabhängig von den Auswirkungen der Unerwartetheit. Zusammengefasst fanden wir keine Hinweise darauf, dass die Sensibilität der Leser für das Vorhandensein eines Konnektivs davon abhängt, wie vorhersehbar der Relationstyp oder der Inhalt ist.

In Kapitel 7 wurde untersucht, ob und wie die **Eigenschaften des Lesers** an sich einen Einfluss darauf haben, ob er sprachliche Signale ausnutzt. Genauer gesagt fragten wir ob Fachwissen die Interpretation von Diskursrelationen und die Nutzung sprachlicher Signale durch den Leser beeinflusst. Wir baten Experten aus den Bereichen Biomedizin und Wirtschaft Konnektive in Texte aus ihrer eigenen Domäne (z.B. biomedizinische Experten, die biomedizinische Forschungsarbeiten lesen) und aus der anderen Domäne (z.B. biomedizinische Experten, die Wirtschaftszeitungen lesen) einzufügen. Diese Konnektive wurden in Relationsklassen eingeteilt, um ihre Genauigkeit bei der Ableitung der Zielrelation zu untersuchen. Biomedizinische Ex-

perten waren bei der Ableitung von Relationen in biomedizinischen Texten erfolgreicher als Wirtschaftsexperten, was zeigt, dass die Leser ihr Fachwissen nutzen, um Diskursrelationen abzuleiten. Bei der Interpretation von Relationen in den Wirtschaftszeitungen wurde kein solcher Unterschied festgestellt, wahrscheinlich weil diese sich an ein breiteres Publikum richten als biomedizinische Forschungsarbeiten. Wenn die Relation nicht auf der Grundlage des vorhandenen Wissens abgeleitet werden kann, nehmen die Leser unspezifizierte Interpretationen vor. Darüber hinaus zeigen wir, dass Leser sprachliche Signale für Diskursrelationen nutzen, aber dass solche Hinweise manchmal Fachwissen erfordern. Um zum Beispiel zu verstehen, dass zwei Begriffe Antonyme sind, müssen die Leser zunächst wissen, was diese Begriffe bedeuten. Doch selbst wenn diese Signale nur allgemeines Wissen erfordern, nutzten Leser mit geringen Kenntnissen sie nicht immer. Dies deutet darauf hin, dass nicht-konnektive Signale vom Leser nur genutzt werden um bereits bestehende Interpretationen von Diskursrelationen zu bestätigen.

Die Untersuchungen in dieser Dissertation liefern wertvolle Einblicke in verschiedene Theorien der Sprachverarbeitung. Zunächst einmal zeigen die Ergebnisse in Bezug auf die Diskursverarbeitung, dass Leser sowohl sprachliche (Konnektive und nichtkonnektive Signale) als auch außersprachliche Informationsquellen (Hintergrundwissen) nutzen, um Kohärenz herzustellen. Wir zeigen auch, dass die Sensibilität der Leser für Signale von Diskursrelationen von den Eigenschaften des Signals, der Relation und auch des Lesers selbst abhängt. So ist die Wirkung eines sprachlichen Signals stärker, wenn es auffälliger und informativer ist, die Relation nicht kausal ist und der Leser sich auf vorhandenes Wissen verlassen kann, um die Bedeutung des Signals zu bestätigen. Zweitens leisten die hier vorgestellten Untersuchungen einen Beitrag zu informationstheoretischen Konzepten der Sprachverarbeitung, indem sie zeigen, dass die Unerwartetheit des Inhalts und der Diskursstruktur die Verarbeitungsschwierigkeiten beeinflussen, aber nicht die erleichternde Wirkung des Konnektivs erklären. Drittens leistet die vorliegende Dissertation einen Beitrag zur Forschung über statistisches Lernen, indem sie aufzeigt, dass Leser sich der Korrelationen zwischen syntaktischer Struktur und Bedeutung auf Diskursebene bewusst sind. Viertens zeigt die Arbeit im Hinblick auf die Forschung zu individuellen Unterschieden, dass die Interpretation von Diskursrelationen und die Verwendung von sprachlichen Signalen vom Hintergrundwissen der Leser abhängt. Fünftens erweitern wir die sprachübergreifende Forschung zur Sprachverarbeitung und stellen fest, dass es sprachübergreifende Ähnlichkeiten bei der Verwendung von Konnektiven zur Verarbeitung von Diskursrelationen gibt. Schliesslich unterstreicht diese Dissertation auch die Bedeutung konvergenter Evi-

denz, indem sie zeigt, dass die Kombination von verschiedenen wissenschaftlichen Methoden die Theoriebildung erleichtern kann.

Acknowledgments

During this PhD, I have been blessed with so many people who have helped me reach this point and created many memorable moments along the way.

Firstly, Vera, thank you for guiding me these last four years. Your ideas, advice and (incredibly detailed!) feedback contributed tremendously to the development of this work and my research skills. Thank you for being so involved in every single project, despite supervising so many people. Your keen intellect, hard work and academic curiosity are truly inspiring. Thank you also for giving me the freedom to wander off on side projects and to work from anywhere but the office so often.

Merel, I am very grateful that our paths crossed: You have not only been a great collaborator, excellent supervisor and wonderful travel companion, but also an inspiring mentor and a dear friend. Thank you for teaching me everything I needed to know to complete this dissertation. Your knowledge of the field, the high quality and pace of your work and your project management skills are incredible. Most importantly, you always remain positive, kind and encouraging. Thank you for believing in me throughout the ups and downs of my research and personal life.

Over the course of my PhD, I have had the pleasure of collaborating with many people. Ted, thank you for sparking my interest in discourse processing research and creating so many opportunities for me to pursue this line of research. Thank you for reminding me to stand on the shoulders of giants. Frances, I enjoyed our discussions on the difficulties of discourse annotation and evaluation (and other aspects of life). Freya, it was lovely working with you, Sara and Manfred on the CLC project. Thank you to Hannah Rohde, for hosting me in Edinburgh. A big thank you to everyone who worked on the eye-tracking experiment: Benedek, Sinead, Jiong, Markéta, Domenik, Brianna and Sophia. It was truly a team effort. Aria-Ray, thank you for going over so many versions of my materials and your efforts in the Pidgin project that took forever. The Hiwis responsible for LingoTurk over the years, who made working with a buggy system so smooth, also deserve a shout out. Vini and Clare, thank you for the many inspiring meetings and sessions we've had with DiscoMatiX.

I am very grateful to have done my PhD in such a large and warm lab: Thank you to all my fellow PhD-students, post-docs and visiting researchers for sharing struggles and celebrating successes. Thank you also for the many group lunches and dinners, board game nights and Christmas activities. A big thank you to Khanda, for helping us deal with German bureaucracy and always being patient and kind. Thank you to all the PLEAD members (Katja, Miriam, Marjolein, Pratik, Merel, Emilia, Rita, Iza,

Jia, Sebastian, Wei, Christina, Isabell, Jack, Laura, Hannah and many others) for providing me with feedback on my designs and analyses, presenting their work and learning about statistics together.

Thank you to SFB1102, personified in the Board and the Coordination Team, for supporting and funding my research - including the Small Grant Scheme and my many many travels - and all the workshops I wanted to organize. A special thank you to Vera, Matt and Jet for reviewing this dissertation and their inspiring work and feedback, as well as to Ingo, Stefania and Frances for joining my defense committee. Thank you my various PhD support groups for checking up on my progress and sometimes delaying that progress with welcome distractions. Thanks to Cito for welcoming me for my internship: I had a great time and learned so much about all the valuable work (and research) that is done outside of academia. I am excited to contribute to this work in the future.

I want to thank all my friends in Saarbrücken for making me feel at home here. Thanks to my flatmates (especially Daria) for ensuring I actually spoke some German during my time here, and for the many cups of tea, board games and flammkuchen dinners. Thank you to Marjolein, Iza, Mayank, Sasha, Jia, Pratik, Anu, Miriam and Rita for our many sports-related (cycling to France, hiking, running, sledging, ice-skating, spikeballing, swimming, bouldering, rollerskating) or food-related (dinners, picnics and BBQs at Staden, visits to the Weihnachtsmarkt) activities. And of course, for celebrating Sinterklaas with me. I'm looking forward to seeing you again, in the Netherlands, Saarbrücken or other exciting places.

To my friends back home in the Netherlands: thank you for the get-togethers in the Netherlands, the summer vacations all across Europe and weekend getaways around Maastricht and in Saarbrücken. Thank you, Louise in particular, for the many real-life and virtual teas and movie nights. Thank you Roelieke and Marjolein for friendships that survive not seeing each other for a long time.

A special thanks goes out to my (extended) family, who have supported me in following my dreams - even when they did not always understand them - by providing a warm and welcoming place whenever I returned home. Papa en mama, Els en Adeel, Jannine en Geert-Pieter, Jurre en Nore en Shiloh, dank jullie wel dat ik altijd bij jullie thuis mag komen.

Finally, Bertram, thank you for always being there for me, despite the distance. Thank you for listening to my many worries and distracting me with delicious food and (camping) trips. Thank you for always encouraging me to be kind to others and myself. Thank you for reminding me what is important in life. Dank je wel!

Contents

Part I

Theoretical framework

Chapter 1

Introduction

Language is a central aspect of every day life. We use language to tell jokes to our loved ones, listen to a friend's ideas, read a news item on recent elections and reply to an email from a coworker. These processes often seem effortless, but producing and understanding language is an intricate process, which is an ability that is unique to humans (Pinker & Jackendoff, 2005).[1] Thus, understanding the processes and mechanisms involved in language production and comprehension sheds more light on human cognition.

A key element of communication is the ability to comprehend longer excerpts of text, whether written or spoken. Such a text is often referred to as a discourse. This dissertation aims to contribute to understanding the processes involved in discourse comprehension. More specifically, the research presented here focuses on how readers make sense of a text: How do comprehenders know that a discourse is coherent?

Consider the Example in (1), taken from *Alice in Wonderland* by Lewis Carroll (1993, p. 18).

(1) The rabbit-hole went straight on like a tunnel for some way, and then dipped suddenly down, so suddenly that Alice had not a moment to think about stopping herself before she found herself falling down a very deep well.

[1] Large language models (LLMs), like the GPT family (cf. Radford et al., 2018), show remarkable linguistic competence, especially regarding the form of language (cf. Mahowald et al., 2024). Interestingly, they perform considerably less well when evaluated on functional linguistic competence (i.e. the use of language, Mahowald et al., 2024), including tasks relating to discourse coherence (Sieker et al., 2023; Yung et al., 2024a), the topic of the present dissertation.

To understand this excerpt, one does not only need to decode the black symbols (i.e. letters) into meaningful words, but also needs to understand how these words are structured to derive the meaning of each sentence and see how this relates to the other sentences and to existing knowledge. Theories of discourse comprehension assume that readers and listeners construct a mental representation of the text (Zwaan & Rapp, 2006; Van den Broek, 2010). In order for comprehension to be successful, such a mental representation needs to be *coherent* (Sanders et al., 1992; Kehler, 2006). That is, the parts of the mental representation should be connected in a meaningful way. One way in which readers do this is by understanding how different parts of a text are related (Sanders et al., 1992). To illustrate, the first two clauses from the excerpt in (1) above, repeated in (2) below, are related in a temporal way: The rabbit-hole first went straight before suddenly going down.

(2) The rabbit-hole went straight on like a tunnel for some way, and then dipped suddenly down (...) (Carroll, 1993, p. 18)

These logical links between segments in a text are referred to as *discourse relations* (Sanders et al., 1992; Zufferey & Degand, 2024). To establish these relations, comprehenders can draw on both linguistic and non-linguistic information (Zwaan & Rapp, 2006; Van den Broek, 2010). For example, the temporal relation illustrated above is signaled by the connective phrase *and then*. However, such a clear lexical signal is not present for the causal relationship that the sudden drop resulted in no time for Alice to think about stopping herself:

(3) (...) and then dipped suddenly down, so suddenly that Alice had not a moment to think about stopping herself (...) (Carroll, 1993, p. 18)

In this case, readers have to infer the relation using other sources of information.

Although various sources of information for inferring discourse relations have been established in the literature (cf. Noordman et al., 1992; Das & Taboada, 2018b), it is still unclear whether readers rely on them to establish coherence and how these various sources of information influence the processing of discourse relations. The studies presented in this dissertation shed more light on this issue and aim to answer the following research question:

Which factors influence readers' sensitivity to linguistic signals of discourse relations?

In order to answer this research question, various goals for the dissertation were formulated, which are presented in the next section. The specific contributions of this thesis to more general theories of language processing are formulated in Section 1.2. An overview on how these goals were achieved is provided in section 1.3, which also serves as a road map for this dissertation. Section 1.4 highlights how the findings of the present dissertation have been disseminated.

1.1 Research goals

Readers' sensitivity to information about the discourse relation may depend on a variety of factors. The research in this dissertation aims to investigate four factors, presented here as separate goals.

1 The first goal of this dissertation is to examine which **characteristics of the linguistic signal** influence readers' use of this information in processing discourse relations. Connectives are prototypical signals of discourse relations, but there are many other linguistic sources of information about the relation. However, it is unclear to what extent readers use these various types of signals when processing discourse relations. In Chapter 3, I therefore review previous literature on connective and non-connective signals of discourse relations, identifying features of these signals that influence their effect on discourse processing and representation. Furthermore, I present an experimental investigation on the role of a non-lexical discourse signal, namely clause structure, in discourse processing and representation in Chapter 6. These findings are contrasted with studies examining the effect of the presence of a connective in Chapters 4 and 5.

2 The second goal of the thesis is to investigate the role of **characteristics of the discourse relation** in readers' sensitivity to linguistic signals of discourse relations. To this end, I examine whether the type of discourse relation as well as the predictability of the discourse relation influence the effect of the connective on discourse processing. Some discourse relations have been argued to be more difficult to process than other relations. In Chapter 4, I compare the facilitating effect of the connective on reading in `result` relations with `concession` and `contrast` relations. Furthermore, readers' sensitivity to the presence of a connective might be influenced by the predictability of the discourse relation. In Chapter 5, I tease apart different ways in which a connective can make the upcoming material more predictable and investigate

whether the facilitating effect of the connective can be attributed to this enhanced prediction.

3 The third goal of this dissertation is to research how **characteristics of the reader** influence their sensitivity to information about discourse relations. More specifically, I examine how differences in background knowledge affect readers' interpretation of discourse relations in Chapter 7. In addition, I investigate their use of linguistic signals of discourse relations and other strategies for inferring discourse relations.

4 The fourth goal is to investigate whether **characteristics of the language** affect how sensitive readers are to lexical signals in the processing of discourse relations. Typological differences between languages could influence how readers process a language, which might also have consequences for processing discourse processing. In Chapter 4, I compare the effect of the connective on processing discourse relations in speakers of English and German.

1.2 Research contributions

The research in this thesis contributes to various theories on language processing. Below, we discuss the main findings in this dissertation and how they relate to existing theories and research.

Theories on processing discourse relations. Regarding models of discourse processing, there are two main contributions. The first is that **readers exploit various sources of information to establish coherence**. On the one hand, readers use *linguistic cues* to construct a coherent mental representation. In Chapters 4 and 5, I find that connectives facilitate the processing of subsequent material. In addition, I show that even a non-lexical linguistic cue, namely clause structure, influences readers expectations about the upcoming discourse structure (Chapter 6). On the other hand, readers draw on *extra-linguistic sources* of information to infer discourse relations. Chapter 7 reveals that readers infer the intended discourse relation more accurately in texts from their domain of expertise, compared to texts from a domain in which they have little expertise. The second contribution for theories on discourse processing is that I show that **readers' sensitivity to signals of discourse relations depends on characteristics of the signal, the discourse relation and**

the reader. Based on a review of the literature, I argue in Chapter 3 that the extent to which readers rely on a linguistic signal of a discourse relation is determined by five characteristics of the signal: informativity, functionality, immutability, lexicality and agreement. Signals that adhere to these features more strongly will influence discourse representation and processing more strongly than those that do not. This is illustrated by the findings in Chapters 4, 5 and 6. I find strong evidence for the effect of connectives on on-line processing in Chapters 4 and 5. However, no evidence was found that clause structure, a non-lexical cue, influences the on-line processing of discourse relations, nor preferences in how the discourse relation is formulated (see Chapter 6). Furthermore, the effect of a signal depends on features of the discourse relation it marks: Chapter 4 shows that connectives facilitate processing more in `concession` relations than in `result` relations. In addition, Chapter 7 reveals that readers' use of linguistic signals for discourse relations depends on characteristics of the reader: domain knowledge is often required to exploit non-connective cues when inferring discourse relations. This also shows that readers need to *combine* information from linguistic and extra-linguistic sources to establish coherence.

Information-theoretic accounts of language processing. Information-theoretic accounts of language processing assume that processing difficulty is proportional to surprisal, the negative log probability of the linguistic material in context (Levy, 2008; Demberg & Keller, 2008; Wilcox et al., 2023). Previous studies have mainly focused on the unexpectedness of lower-level linguistic material, such as phonemes, words and part of speech (Heilbron et al., 2022; Wilcox et al., 2023). Here, I extend this previous research by showing that **the unexpectedness of semantic content and discourse structure influence processing difficulty, but do not explain the facilitating effect of the connective**. This suggests that readers predict upcoming semantic content and discourse structure. On the other hand, Chapter 5 also reveals the limits of an information-theoretic account of language processing. Various measures of unexpectedness did not explain the facilitating effect of the connective on processing difficulty.

Statistical learning. Identifying regularities in the linguistic signal has been argued to play an important role in language acquisition, also referred to as *statistical learning* (cf. Saffran et al., 1996). However, there is little evidence on whether language users track patterns at the discourse level (but see Johnson & Arnold, 2021). Chapter 6 reveals that **readers are aware of correlations between linguistic structure and discourse-level meaning**. To be more precise, the co-occurrence

between clause structures and certain discourse relations influenced readers' off-line expectations for upcoming discourse relations. However, no evidence was found that these predictions influence on-line processing or discourse interpretation, suggesting that statistical learning at the discourse-level has a limited effect on discourse processing.

Individual differences. Individuals have been shown to differ in how they process various aspects of language (cf. Kidd et al., 2018), including at the discourse level (e.g. Scholman et al., 2020). Here, I extend previous research by showing that **reader characteristics influence the interpretation of discourse relations**. More specifically, Chapter 7 shows that a reader's domain knowledge influences the accuracy and type of discourse relation inferences. Furthermore, joint work with Merel Scholman and Vera Demberg that is not part of this dissertation also contributed to this goal. This study focused on individual differences in connective comprehension and showed that vocabulary size, non-verbal IQ, and cognitive reasoning style influence to what extent individuals understand the meaning of a connective (Scholman et al., 2024a).

Cross-linguistic comparison. Coherence is assumed to be a cognitive notion that is important across languages. However, languages differ in many respects and the way coherence is marked linguistically might vary, which could influence processing. Only a few studies have compared discourse processing across languages (Schwab & Liu, 2020; Yi & Koenig, 2021). The research presented in this dissertation reveals that **there are cross-linguistic similarities in the use of connectives for discourse relation processing**. This is explicitly tested in Chapter 4, in which I find no evidence that the effect of the connective on processing is different in English compared to German across two experiments. In addition, research not presented in this dissertation examined discourse marking in Naija, a contact language spoken in Nigeria. This showed similarities in how and when relations are marked in Pidgin compared to English (Marchal et al., 2021b; Scholman et al., submitted).

Converging evidence. Finally, the research in this dissertation highlights that **combining evidence from various methodologies can facilitate theory building**. As has been argued before (Graesser et al., 1994; Sanders & Evers-Vermeul, 2019; Scholman et al., 2022b), different methodologies provide different insights into discourse processing and evidence from various sources should therefore be combined. In the literature review on the effect of relation signals on discourse representation and

processing (Chapter 3), I draw on findings from a large variety of methods, showing that some signals do not consistently affect discourse relation processing across tasks. Furthermore, I combine corpus research with off-line and on-line tasks in Chapter 6, which provides comprehensive insights into the extent to which a signal influences discourse processing. This study shows that the role of non-lexical cues in inferring discourse relations might be limited. Together, these findings suggest that readers' *can* use various sources of information in processing discourse relations, but that this information does not always influence inference making.

1.3 Overview of the dissertation

The dissertation is divided into four parts. **Part I**, which also contains the current chapter, provides the theoretical framework for the remainder of the dissertation. It outlines the research questions and goals for the dissertation, gives an overview of related work, and provides a background for the experimental studies in the subsequent parts.

More specifically, **Chapter 2** elaborates on the role of discourse relations in language processing. It discusses how discourse relations can be defined and categorized, as well as how the processing and representation of discourse relations can be investigated experimentally. The chapter also outlines various theories on how discourse relations are processed.

Chapter 3 discusses when and how discourse relations are signaled, and how such signals affect discourse representation and processing. Five features are identified to distinguish various types of discourse relation signals. Based on a review of previous literature on the effect of signals on discourse representation and processing, I argue that these features influence to what extent readers rely on a signal. This aligns with Research Goal 1.

Following this theoretical framework, four experimental studies are presented. **Part II** of the dissertation focuses on the role of connectives in processing discourse relations.

Chapter 4 examines whether the facilitating effect of the connective on processing depends on language and relation type, addressing Research Goals 2 and 4. It presents the findings from two self-paced reading experiments investigating the interaction of the presence of a connective with language (English vs. German) and with relation type (`result` vs `contrast` and `concession`). The results show that the connective

facilitates reading of the subsequent material (but only in Experiment 2) and that this effect is larger in `concession` relations than in other relations. No such interaction was found between language and the presence of a connective in either experiment, suggesting that the facilitating effect of the connective is not language-dependent.

Chapter 5 investigates whether the facilitating effect of the connective can be attributed to prediction. More specifically, it is hypothesized that readers continuously make predictions about the upcoming discourse relation, semantic content and lexical-syntactic structure and that these predictions facilitate processing. Since the presence of a connective might also influence the predictability of the upcoming material, this could explain the facilitating effect of the connective found in previous studies. I show that the presence of a connective indeed reduces the unexpectedness of the discourse relation and, indirectly, of the semantic content, and that these two factors influence reading times in both a self-paced reading and an eye-tracking-while-reading experiment. However, these effects could not explain the facilitating effect of the connective. This chapter contributes to Research Goal 2.

Beyond connectives, comprehenders might also make use of other sources of information, both linguistic and non-linguistic. **Part III** of the dissertation presents two experimental studies investigating readers' use of non-connective information.

Chapter 6 explores whether readers rely on a non-lexical signal in making discourse relation inferences. A corpus investigation shows that gerund free adjuncts often cooccur with `result` relations. This distribution of relation types is also reflected in readers' off-line expectations about discourse relations, showing that they predict more `result` relations after a gerund free adjunct compared to a full matrix clause. However, these expectations did not facilitate the processing of `result` relations in a self-paced reading task, nor did the co-occurrence between this clause type and `result` relations influence readers' preferences for the formulation of the discourse relation. The findings in this chapter contribute to Research Goal 1.

Chapter 7 investigates the role of domain knowledge in the interpretation of discourse relations. Experts from biomedical sciences and economics inserted connectives in in-domain and out-of-domain texts. High-knowledge readers were found to make more accurate interpretations. In addition, strategies that readers use in the absence of required domain knowledge were explored, showing that low-knowledge readers prefer less precise interpretations. Furthermore, I show that the use of non-connective cues often requires domain knowledge, suggesting that readers need to combine vari-

ous source of information when inferring discourse relations. This chapter addresses Research Goals 1 and 3.

The last part, **Part IV**, relates the findings of the various parts of the dissertation to each other and to earlier research and presents an outlook for further research.

Chapter 8 provides an answer to the overall research question about the factors influencing readers' sensitivity to linguistic signals of discourse relations. It summarizes the findings of the studies in this dissertation and connects them to related work.

Chapter 9 specifies directions for future work and concludes the dissertation.

The chapters are connected by the overarching research question presented at the beginning of this chapter, but they can all be read in isolation.

1.4 Dissemination

The work in this dissertation is (in part) also described in the following peer-reviewed publications:

Marchal, M., Scholman, M.C.J., & Demberg, V. (2022). The effect of domain knowledge on discourse relation inferences: Relation marking and interpretation strategies. *Dialogue & Discourse, 13*(2), 49-78. doi:10.5210/dad.2022.202 (cf. Chapter 7)

Marchal, M., Scholman, M.C.J., & Demberg, V. (2023). How statistical correlations influence discourse-level processing: Clause type as a cue for discourse relations. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *50*(5), 796-807. doi:10.1037/xlm0001270 (cf. Chapter 6)

Marchal, M., Hewett, F.[2], Scholman, M.C.J., Shahmohammadi, S., Stede, M. & Demberg, V. (submitted). The facilitating effect of the connective is dependent on the relation, but not on language. *Discourse Processes.* (cf. Chapter 4)

Marchal, M., Scholman, M.C.J., Sanders, T.J.M., & Demberg, V. (2024). What processing instructions do connectives provide? Modeling the facilitative effect

[2]Equal contribution.

of the connective. In *Proceedings of the 46th Annual Meeting of the Cognitive Science Society* (pp. 3435–3441). (cf. Chapter 5, Pretest and Experiment 1)

Marchal, M., Scholman, M.C.J., Sanders, T.J.M. & Demberg, V. (in prep.). Predicting discourse relations: Understanding the facilitating effect of the connective. (cf. Chapter 5)

Marchal, M., Scholman, M.C.J., Demberg, V. & Sanders, T.J.M. (in prep.). Discourse relation signals in discourse processing and representation: Characteristics of the signal and the reader. (cf. Chapter 3)

The research in this dissertation has also been presented at a variety of conferences. These are listed at the end of each of these chapters.

Although these papers are all co-authored, the conceptualization, methodology, investigation, formal analysis, writing (both original draft and review and editing) as well as visualization of these manuscripts are predominantly my own. An exception to this is Experiment 2 in Chapter 4, which was conceptualized and conducted by Freya Hewett, Sarah Shahmohammadi and Manfred Stede. Still, I was closely involved in the analysis of the collected data as well as the conceptualization and investigation of Experiment 1 and the writing of the manuscript.

Parts of research that was carried out during the dissertation but is not included here can be found in the following publications:

Marchal, M., Scholman, M.C.J., & Demberg, V. (2021). Semi-automatic discourse annotation in a low-resource language: Developing a connective lexicon for Nigerian Pidgin. *Proceedings of the Second Workshop on Computational Approaches to Discourse (CODI)*, 84-94.

Marchal, M., Scholman, M.C.J., Yung, F. & Demberg, V. (2022). Establishing annotation quality in multi-label annotations. *Proceedings of the 29th International Conference on Computational Linguistics* (COLING), 3659-3668.

Scholman, M.C.J., **Marchal, M.** & Demberg, V. (2024). Connective comprehension in adults: The influence of connective features and individual differences. *Discourse Processes, 61*(8), 381–403.

Scholman, M.C.J., **Marchal, M.**, Brown, A.R. & Demberg, V. (in prep.). DiscoNaija: A discourse-annotated parallel corpus of English and Nigerian Pidgin. *Language Resources and Evaluation.*

Chapter 2

Discourse relations

Successful comprehension requires constructing a coherent mental representation of a text (Zwaan & Rapp, 2006; Van den Broek, 2010). An important aspect of this representation are discourse relations. Discourse relations are the logical links between segments in a text (Sanders et al., 1992; Zufferey & Degand, 2024), and have also been referred to as *coherence relations* (Hobbs, 1979; Sanders et al., 1992) or *rhetorical relations* (Asher & Lascarides, 2003; Sporleder & Lascarides, 2008). As we will see below, they form a crucial aspect of discourse representation. This chapter will discuss what discourse relations are and what role they play in establishing discourse coherence.

Furthermore, this dissertation aims to investigate how discourse relations influence processing. Discourse annotations provide insights both into the distribution of discourse relation types and the interpretation of these relations, which is discussed in Section 2.3. Subsequently, Section 2.4 provides background on and the rationale for the experimental research methods adopted in this dissertation. The chapter ends with a review of previous literature on the role of (different types of) discourse relations in representation and processing. Gaps in the literature, which will be addressed in the following chapters, are identified.

2.1 Establishing a coherent mental representation

When readers process a text, they do not only try to understand the meaning of each individual sentence, they also aim to understand how the information in different parts of the text is related to each other (Hobbs, 1979; Sanders et al., 1992). In other words, readers constantly aim to establish coherence (Hobbs, 1979). Consider the following example.

(4) Mary is cycling to work. She loves pizza.

At first sight, these sentences are not related. Nevertheless, readers will try to find a link between the two propositions, maybe that Mary has eaten a lot of pizza lately, because she loves it, and is now trying to compensate for these calories by cycling to work. This "desire for coherence" thus goes beyond congruity between sentences in their topic (Hobbs, 1979, p. 67), also referred to as *content coherence* (Sanders et al., 1992). Readers also tend to infer *how* different segments of a discourse are related (Hobbs, 1979; Sanders et al., 1992). This is referred to as *relational coherence* (Sanders et al., 1992).

Discourse coherence thus refers to the connectivity between the different aspects of a mental representation of a text and is a crucial aspect of successful comprehension (Sanders et al., 1992). It is a cognitive notion, rather than a linguistic one: discourse coherence is grounded in the mental representation readers construct of a text and is also established in the absence of linguistic markers (Sanders et al., 1992). As such, it is different from *cohesion*, which refers to the links between linguistic elements whose interpretation is dependent on that of another element, such as pronouns (Halliday & Hasan, 1976). Cohesion is grounded in the linguistic realization of the connections in a text, whereas coherence can only be established in the mind of the reader (Sanders et al., 1992; Zufferey & Degand, 2024). The present chapter focuses on the cognitive level of coherence. Linguistic signals of coherence will be discussed in the next chapter.

2.2 Defining discourse relations

A discourse relation refers to the relationship between segments in a discourse (Sanders et al., 1992; Zufferey & Degand, 2024). It is the aspect of meaning that goes beyond the meaning of the segments in isolation (Sanders et al., 1992). This relational surplus is what makes a discourse coherent and the mental representation more than the sum of the meaning of the individual segments (Sanders et al., 1992; Zufferey & Degand,

2024). Below, I will describe how the segments of a discourse relation as well as the different types of relations can be defined.

2.2.1 Segmenting the discourse

If a discourse relation refers to the link between discourse segments, what constitutes such a segment? The smallest, and most common, segment of a discourse relation is traditionally considered to be an independent clause (Mann & Thompson, 1988; Prasad et al., 2008; Sanders et al., 1992). This is illustrated in Example 5 below, with the two segments delineated with square brackets.

(5) [Mary is planning on going to Italy.]$_1$ [She loves pizza.]$_2$

These clauses can be separate sentences, as in the Example above, or two clauses within the same sentences, as in Example 6. Note that lexical markers of the discourse relation are not included in the argument.[1]

(6) [Mary is planning on going to Italy,] because [she loves pizza.]

A strict definition of the clause as the smallest unit has been relaxed in later work (Taboada & Mann, 2006; Webber et al., 2019), because such a syntactic criterion does not generalize well across languages (Taboada & Mann, 2006; Stede et al., 2018; Marchal et al., 2021b) and modalities (Hoek et al., 2018). Ideally, the segments of a discourse relations are idea units, but these are not always expressed in clauses (Hoek et al., 2018). For example, the second segment of the Example 7 is not an independent clause, but does constitute an idea unit.

(7) [Mary is planning on going to Italy,]$_1$ because of [her love for pizza.]$_2$

Discourse segmentation is usually done independently from determining the sense to prevent circularity (Taboada & Mann, 2006). However, the criticism on a syntactic definition of a discourse segment outlined above has led researchers to focus more on the propositional content of the segment when defining what constitutes a segment. For example, Stede et al. (2018) argue that the segments should represent abstract concepts, which can, but need not be, expressed in clauses. However, annotators might disagree on what such an abstract concept is. Hoek et al. (2018) propose using paraphrase tests to determine whether segments could in principal be a discourse

[1] We follow the Penn Discourse Treebank in this respect, as discussed below in Section 2.2.2.

relation (i.e. contain propositional meaning). To illustrate, the second argument in Example 7 above can be paraphrased as Example 6 above. This is not the case for the second argument in Example 8.

(8) [Mary is planning on going to Italy,]$_1$ because of [pizza.]$_2$

Hoek et al. (2018) argue that (parts of) predicates, rather than clauses, constitute the smallest structural unit for a discourse segment. Following this line of reasoning, Hoek et al. (2018) suggest that also embedded clauses, such as relative clauses, can be part of a discourse relation (see Hoek et al., 2021a, for empirical evidence that readers update discourse expectations based on information from relative clauses). In Chapter 6, I contribute to this line of research by investigating a different type of predicate, that is also not an independent clause: gerund free adjuncts.

Segments can also consist of larger units, such as multiple sentences, full paragraphs or even chapters (Sanders et al., 1992; Mann & Thompson, 1988). Frameworks differ in how these more global discourse relations are indicated. Some frameworks prefer hierarchical organizations (Mann & Thompson, 1988, see also Hoek et al., 2018), whereas others do not specify that segments can consist of discourse relations themselves, resulting in a flat discourse structure with more local annotations (Prasad et al., 2008). In these cases, a useful guideline is the minimality principle, which states that only the clauses which are minimally required and sufficient for interpreting the relation should be included (Prasad et al., 2008).

Another issue in discourse segmentation is whether all clauses or predicates are part of the discourse structure. Hoek et al. (2018) argue that this completeness constraint (Mann & Thompson, 1988) applies to all elements that are part of the propositional content. Annotation efforts thus usually include all segments that occur in the discourse, regardless of whether they are linguistically signaled or not (though of course the definition of a segment is framework-dependent). However, discourse segmentation sometimes favors explicitly marked segments. For example, Carlson & Marcu (2001) include elliptical clauses only if they contain discourse marking. Similarly, in their multilingual discourse-annotated corpus, Zeyrek et al. (2020) only include intra-sentential discourse relations when they are signaled with a connective. This should be taken into account when comparing distributions of relation marking across corpora.

2.2.2 Categorizing discourse relations

There are various ways in which segments can be related. For instance, the relation between segments can be additive, with the content of one segment elaborating on that of the previous one, as in (9a). In causal relations, illustrated in (9b), one segment provides the consequence of the cause in another segment. If the events in the segments are not causally related, but can be ordered temporally, as in (9c), they are referred to as temporal relations. In adversative relations, there is a contrast between (the expectations raised by) the content in one segment and the other segment. Finally, in conditional relations, one segment provides a condition for the content of another segment, see (9e).

(9) a. Mary is a nurse. She works at the local hospital. `elaboration`

b. Mary hadn't had time for lunch, so she was very hungry. `cause`

c. When she came home from work, she immediately popped a pizza into the oven. `succession`

d. Mary loves Italian food, but her sister does not. `contrast`

e. If she hadn't been so hungry, she would have taken the time to cook her favorite dish, lasagna. `condition`

In addition to the distinctions above, many other and more specific relation types can be distinguished. Indeed, many different ways to classify discourse relations have been proposed (e.g. Hobbs, 1979; Kehler, 2002; Asher & Lascarides, 2003). Although these taxonomies, also referred to as frameworks, usually distinguish the categories outlined above in some way, they diverge on how these categories should be labelled and which other (sub-)relation types exist. In this dissertation, the classification used in the Penn Discourse Treebank (PDTB, Prasad et al., 2008; Webber et al., 2019) is adopted for discourse annotation and corpus research. However, two other frameworks are also relevant to the work presented here. Rhetorical Structure Theory (RST, Mann & Thompson, 1988; Carlson & Marcu, 2001) has been used in the largest annotation effort of non-connective signals for discourse relations, discussed in the next chapter. The Cognitive Approach to Coherence Relations (CCR, Sanders et al., 1992; Hoek et al., 2019a) is based on cognitive distinctions in discourse relation classification and provides useful dimensions for discussing the effect of discourse relations on processing, as well as combining different frameworks (Sanders et al., 2021). For this reason, these three frameworks will be introduced shortly below. For a comprehensive overview, the interested reader is referred to Zufferey & Degand (2024).

The Penn Discourse Treebank framework stems from an effort of annotating the 1 million word Wall Street Journal Corpus in the Penn Treebank with discourse relations (Prasad et al., 2008). The most recent version, PDTB-3, consists of about 40k relation annotations (Webber et al., 2019). Since its introduction, the framework has also been used in multiple other corpora, including BioDRB (Prasad et al., 2011), TED-MDB (Zeyrek et al., 2020) and the Disco-GEM corpora (Scholman et al., 2022c; Yung et al., 2024b). The starting point of the Penn Discourse Treebank are lexical signals of discourse relations, such as connectives. When these connectives are not present in the original text (i.e. the relation is implicit), annotators insert a connective to determine the corresponding label. Segments are referred to as arguments. In implicit relations, the textual order determines what the first (Arg1) and second argument (Arg2) are. For explicit relations, the second argument is always the one containing the connective. Unless otherwise stated, we will adopt this terminology in this dissertation.

The relational inventory of the PDTB is organized hierarchically. At the top level, four categories, called senses, are distinguished: temporal, contingency, comparison and expansion. These categories are into divided into 22 senses at Level-2. The third level distinguishes the direction of the relation. To illustrate, the relation in Example (9b) would be labeled as `Contingency.Cause.Reason`: a `contingency` relation at the top level, `cause` at the second level and `reason` at the third level. All labels in the PDTB-3 inventory can be found in the Appendix A. Throughout this dissertation, we will refer to relation types mainly using their PDTB-3 Level-3 label. Not all relational classes make a distinction at the third level (e.g. `synchronous`), in which case we will refer to the Level-2 label.

Rhetorical Structure Theory was originally developed by Mann and Thompson (1988), but has been developed further in later work (see Taboada & Mann, 2006, for a discussion). Originally, the relational inventory consisted of 24 relations (Mann & Thompson, 1988), but this has been expanded to 78 relations (Carlson & Marcu, 2001). RST is inherently hierarchical (unlike CCR and PDTB, but similar to Segmented Discourse Representation Theory, Asher & Lascarides, 2003). The smallest possible segment, usually an independent clause, is called an *elementary discourse unit* (Carlson & Marcu, 2001; Taboada & Mann, 2006). Units are divided into a nucleus, containing the most important information, and a satellite, which is less essential. Most relations contain both a nucleus and a satellite, but they can also be multinuclear, in that both units are a nucleus. Units are combined into relations in a recursive manner, such that two independent clauses form a relation, which itself is a

unit that can be combined with another unit in a new relation. As such, the framework combines local and global levels of coherence. There are several corpora that have been annotated in the RST framework (e.g. Carlson & Marcu, 2001; Hewett, 2023), the most extensive being the RST-DT corpus (Carlson & Marcu, 2001) with about 20k relations. This corpus has been updated with a layer of annotations on lexical and non-lexical signals of discourse relations in the RST Signalling Corpus (Das & Taboada, 2018a), which will be discussed in more detail in the next chapter.

The RST and PDTB frameworks aim to classify all discourse relation types in a hierarchical framework. The Cognitive Approach to Coherence relations (CCR, Sanders et al., 1992; Hoek et al., 2019a) classifies types of discourse relations based on features that have been shown to reflect cognitive categories. Although it can also be used for discourse annotation (Scholman et al., 2016; Hoek et al., 2019a), its primary goal is to provide a cognitively plausible classification of discourse relations (Sanders et al., 1992). In fact, the original framework explicitly does not aim to be exhaustive (Sanders et al., 1992), contrary to the RST and PDTB frameworks. CCR classifies discourse relations using several basic primitives, which apply to all discourse relations (Sanders et al., 1992). They are *basic operation* (whether the relation is causal or additive), *source of coherence* (whether the relation exists between the propositions (i.e. semantic) or the illocutions (i.e. pragmatic) of the segments), *order of the segments* and *polarity*. The latter two refer to whether the order is basic (when the first segment in the text refers to P in P & Q or P $\rightarrow$ Q) or non-basic (when the first segment in the text refers to Q) and whether the conjunction or implication relation holds between the positive version of the segments or their negative counterparts. Further distinctions, which apply to only a subclass of the relations, have been proposed in later work, including *temporality*, *disjunction*, *volitionality*, *purpose* and *directness* (Evers-Vermeul et al., 2017; Hoek et al., 2019a). There are various sorts of empirical findings showing that the primitives affect annotation accuracy, language acquisition, representation and on-line processing (cf. Sanders et al., 1992; Evers-Vermeul et al., 2017; Hoek et al., 2019a; Sanders et al., 2021; Scholman, 2019). Some of these aspects will be discussed in the following sections, in which I will refer to these distinctions.

2.3 Annotating discourse relations

Discourse annotations are an important method for research on how readers process discourse relations. They can provide insights into how readers interpret the relations

between segments of a discourse (cf. Rohde et al., 2016; Scholman & Demberg, 2017b, see also Chapter 7) as well as which discourse relations readers expect (cf. Kehler et al., 2008; Scholman et al., 2020, see also Chapters 5 and 6). In addition, discourse-annotated corpora can be used to investigate when and how discourse relations are signaled (see Chapter 3), which can inform research on how these co-occurrences in natural language influence processing (see Chapter 6). For the interested reader, I therefore first review how the type of discourse relation can be determined. Then, I outline some challenges in discourse annotation, showing that discourse relation annotation partly depends on the background knowledge of the annotator, as empirically investigated in Chapter 7. This is followed by a discussion of the advantages of crowd-sourcing discourse annotations, a method also used in Chapter 7, and how discourse annotation efforts (such as in Chapters 5 and 6) can be evaluated.

2.3.1 Determining the relation type

Given a fixed set of discourse relation categories, how can the type of relation be determined? The most obvious cue for the discourse relation is linguistic marking such as connectives (Knott & Dale, 1994; Prasad et al., 2008).[2] For example, the connective *nevertheless* in (10) indicates the `concession` relation.

(10) Mary ate a whole pizza. Nevertheless, she was hungry. `concession`

However, there is not always a one-to-one correspondence between connectives and labels. To illustrate, *but* can be used in both `contrast` as well as `concession` relations (Asr & Demberg, 2020), as shown in (11) and (12).

(11) Mary was craving something sweet, but she ordered a pizza. `concession`

(12) Mary was craving something sweet, but Joe wanted a salty snack. `contrast`

One option is to replace the existing connective with a more specific one, also referred to as the *substitution test* (Scholman et al., 2016). When *but* is substituted with the `concession` marker *nevertheless*, the meaning of (11), but not of (12), remains the same. Conversely, replacing *but* with *in contrast* works for (12) and not for (11). Connectives can also be inserted when no linguistic marker is present, (Sanders et al., 1992; Scholman & Demberg, 2017a), which is done in the PDTB annotation framework (Webber et al., 2019).

[2]We will return to connectives and other cues for discourse relations in the next chapter.

Connectives might not always suffice to distinguish between relational categories. For example, English does not have causal connectives that distinguish between objective and subjective causality (i.e. whether the causal relationship is established in the real world or in the mind of the speaker), as illustrated in (13) and (14) below.[3] In these cases, paraphrase tests can be used to determine the relation (Sanders et al., 1992; Sanders, 1997). They require the annotator to restate the information in the segments in an (often predefined) way that distinguishes different relations and choose which paraphrase best reflects the relation between the segments. To illustrate, Example 13 below can be paraphrased as 'S1 causes the fact that S2'. This does not work for (14), which is best paraphrased as 'S1 causes the speaker's claim that S2' (Sanders, 1997).

(13) Mary had pizza because she loves Italian food.
- ✓ The fact that Mary loves Italian food causes the fact that Mary had pizza.
- ✗ The fact that Mary loves Italian food causes the speaker's claim that Mary had pizza.

(14) Mary had pizza because there is an empty box on the table.
- ✗ The fact that there is an empty box on the table causes the fact that Mary had pizza.
- ✓ The fact that there is an empty box on the table causes the speaker's claim that Mary had pizza.

Paraphrases can also be done in the form of questions. Pyatkin et al. (2020) provide question templates to distinguish between various PDTB relations. For example, the `reason` relation in (13) can be paraphrased as the following question-answer pair:

(15) *What is the reason that* Mary had pizza? She loves Italian food.

2.3.2 Challenges in discourse relation annotation

Annotations are often done by experts using extensive manuals (Carlson & Marcu, 2001; Webber et al., 2019). Nevertheless, even trained annotators do not always agree on which relation holds between the segments (Spooren & Degand, 2010). This raises questions about the reliability and validity of the annotation (Spooren & Degand,

[3]Subjective and objective causality has also been referred to as the distinction between *causal* vs. *diagnostic* relations (Traxler et al., 1997), semantic vs. pragmatic (Sanders et al., 1992) and content vs. epistemic (Sweetser, 1990).

2010; Crible & Degand, 2019). Validity refers to whether the assigned labels reflect the theoretical categories the annotated instances belong to (Artstein & Poesio, 2008; Van Enschot et al., 2024). If these conceptual constructs are not well-defined (low construct validity), only apply to specific situations (low external validity) or are confounded with other factors (low internal validity), agreement will be low (Crible & Degand, 2019; Van Enschot et al., 2024). Reliability refers to whether the same label is assigned consistently, either across instances belonging to the same category, or across different annotators (Van Enschot et al., 2024).

The extent to which annotators agree on discourse relation annotations is influenced by various factors. Agreement is higher when relations are signaled explicitly using connectives (Miltsakaki et al., 2004; Demberg et al., 2019; Hoek et al., 2021c; Kishimoto et al., 2018). Secondly, annotators often agree on some categories, but not on others (Spooren & Degand, 2010; Crible & Degand, 2019). For instance, subjectivity is notoriously difficult to annotate (Spooren & Degand, 2010). Similarly, `contrast` and `concession` are often confused (Sanders et al., 1992; Robaldo & Miltsakaki, 2014; Demberg et al., 2019).

Discourse annotation depends on the interpretation of the annotator (Spooren & Degand, 2010). The annotator has to infer which relation is intended by the author, since not all the information required to determine the relation type might be contained in the text. This suggests that annotation partly depends on the background knowledge of the annotator. Indeed, in Chapter 7, I show that annotators that are experts with respect to the domain of the text achieve higher accuracy than those that are not.

Another challenge in discourse relation annotation is that multiple relations can hold between the segments (Scholman & Demberg, 2017a). Even in the presence of a discourse adverbial, Rohde et al. (2016) show that often another relation can be inferred. For example, both a connective signaling `result` as well as a discourse adverbial signaling `chosen-alternative` can be inserted in the example below.

(16) Mary didn't have time to make lasagna. {so, instead} She made pizza.

Similarly, Scholman & Demberg (2017a) find that `instantiation` and `specification` relations are often interpreted as causal (see also Demberg et al., 2019). A follow-up study (Scholman, 2019) revealed that these can be partly attributed to individual differences in processing depth. However, traditionally, corpus annotations are done by a single coder, often providing only a single relation sense.

2.3.3 Crowd-sourcing discourse annotations

Annotations by a single trained expert are costly and often yield only a single interpretation. Crowd-sourcing methods have helped to alleviate annotation costs (Pyatkin et al., 2023; Scholman et al., 2022d), while also providing richer annotations. Although crowd-workers are not linguistic experts, some of the methods described above also work well in a crowd-sourced setting, even with little training (cf. Scholman et al., 2022d). For example, crowd-sourcing annotations using a step-wise approach with substitution and paraphrase tests (Scholman et al., 2016), connective insertion tasks (Scholman & Demberg, 2017a; Yung et al., 2019) as well as question paraphrases (Pyatkin et al., 2020) have proven to be reliable ways to obtain discourse annotations (see Pyatkin et al., 2023, for a comparison of insertion and question paraphrase methods). Connective insertion tasks have been used extensively in research regarding relation interpretation (Rohde et al., 2016; Scholman & Demberg, 2017b) and to construct corpora with implicit discourse relation annotations (Scholman et al., 2022c; Yung et al., 2024b). This method is also used in Chapter 7 to examine the role of domain knowledge in discourse relation inferences. Furthermore, crowd-sourcing allows for obtaining multiple annotations per instance. These multi-label instances also proved useful in automatic discourse relation classification (Yung et al., 2022; Pyatkin et al., 2023). Note, however, that some sort of quality check is important to prevent low-quality annotations (see Scholman et al., 2022d, for a discussion on training and selecting crowd-workers).

2.3.4 Evaluating discourse annotations

In order to obtain insights from discourse relation annotations, it is important that annotation efforts are reliable. Agreement metrics provide information about the reliability of the annotation. Percentage agreement reflects the percentage of instances where annotators provide the same label. However, there are two issues with this measure. Firstly, this raw measure does not take into account that the overlap in labels could also occur due to chance, which is necessary for making the metric generalizable. Researchers should therefore use a chance-corrected metric of agreement (for an overview, see Artstein & Poesio, 2008). A widely used metric in discourse annotation is Cohen's kappa (Cohen, 1960), which reflects the proportion of agreement beyond chance. In this metric, chance is defined as the probability that an individual annotator assigns an item to a certain category. However, this metric does not capture agreement well when the distribution across the categories is skewed, as

is often the case in discourse relation annotation (Spooren & Degand, 2010; Hoek & Scholman, 2017). Hoek & Scholman (2017) therefore argue that AC_1 (Gwet, 2008) is a more informative metric for agreement in discourse annotation. AC_1 assumes that only some of the assignments are random. Contrary to Cohen's kappa, it takes the total probability of the occurrence of a category (rather than that of an individual coder's bias) and caps chance agreement at 0.5. As a result, chance agreement correlates positively with task difficulty and the metric is less biased by categories that occur only rarely (for a more extensive discussion and comparison of Cohen's kappa and AC_1 in the context of discourse annotation, see Hoek & Scholman, 2017).

Secondly, percentage agreement, but also other agreement measures, can be misleading when multiple labels are assumed to hold. The possibility of multiple interpretations raises the question whether a label that does not match that of the other annotator is truly wrong. In addition, how agreement can be evaluated when multiple labels are provided is a complex issue. One option is to assume agreement for those instances where there is an overlap in at least one of the labels and take this intersecting label as the final label (cf. Crible & Degand, 2019). However, this inflates the probability of achieving agreement based on chance alone, which is not accounted for when subsequently calculating Cohen's kappa (Marchal et al., 2022c). Another option is to consider the proportion of categories also assigned by the other annotator or their weighted average, i.e. precision, recall and F1. However, like percentage agreement, these methods usually do not take into account chance agreement. In a paper not included in this dissertation, I propose a bootstrapping method to account for chance performance on these metrices. This metric provides a more reliable and generalizable way to compare performance on multi-label annotations (Marchal et al., 2022c). In this dissertation, both Cohen's kappa and Gwet's AC_1 will be reported when evaluating annotation efforts with a single label (see Chapters 5 and 6). Annotations with multiple labels (cf. Chapter 6) are evaluated with the measure proposed in Marchal et al. (2022c).

2.4 Experimental methods in discourse research

Discourse annotated corpora can provide empirical insights into the occurrence of various discourse relation types, but the representation and processing of discourse relations can also be investigated experimentally (i.e. by manipulating factors that are hypothesized to influence the representation or processing of discourse relations). Before reviewing prior work on the representation and processing of discourse rela-

tions in Section 2.5, I will first discuss the experimental methods adopted in this dissertation - which have also been used in many previous studies - to provide some background on the described findings. Specifically, I explain what the measured outcomes are hypothesized to reflect and why these methods have been chosen. Note that, unless stated otherwise, all studies in this dissertation were approved by the Deutsche Gesellschaft für Sprachwissenschaft Ethics Committee and informed consent was obtained from all participants.

When designing a study, three factors need to be considered. First of all, the selected method should be valid: it should measure what it is intended to measure. Secondly, the obtained data should be reliable. The data should not be influenced by factors that the researcher is not interested in, as this reduces replicability. Finally, the selected method should be feasible. These three factors usually present a trade-off. Experiments that are easier to conduct are usually less valid. Similarly, high ecological validity often introduces unwanted confounds, reducing the reliability of the results. Below, I will discuss how these factors informed the methodological choices for specific research methods in this dissertation. Following previous work, we will distinguish between off-line tasks and on-line tasks. Off-line tasks tap into the discourse representation, i.e. the product of reading a text. On-line tasks, on the other hand, assess the time-course of comprehension during reading, i.e. the process of reading a text. Both types of tasks are adopted in this dissertation, as they provide complementary insights into discourse processing, as discussed in Section 2.4.3 below.

2.4.1 Investigating discourse representation

To investigate the **interpretation of discourse relations**, we use a *connective insertion task*. In this task, participants are provided with a list of connectives and asked to select the one that best fits the discourse relation (Scholman & Demberg, 2017a). We adopt a two-step approach, in which the free insertion in the first step determines the list of connectives from which a selection can be made in the second step connectives (Yung et al., 2019). This prevents constraining the participants interpretation by the provided list of connectives, while allowing for selection from a large range of connectives that are also relation-specific. In addition to its use in obtaining discourse relation annotations (Scholman et al., 2022c; Yung et al., 2024b), connective insertion tasks have been used to experimentally investigate the interpretation of discourse relations (Rohde et al., 2016; Scholman & Demberg, 2017a,b; Crible & Demberg, 2020) as well as of connectives (Cain & Nash, 2011; Wetzel et al., 2020; Xiao et al., 2021).

To investigate readers' **expectations about discourse relations**, *sentence continuation tasks* were used. This method is a variant of the cloze task, which has traditionally been used to assess the predictability of an upcoming word (cf. Staub et al., 2015). It has also been widely used for examining discourse expectations, such as about referential form and resolution (Kehler et al., 2008; Johnson & Arnold, 2021; Demberg et al., 2023; Rosa & Arnold, 2017) as well as about discourse relations (Kehler et al., 2008; Bott & Solstad, 2021; Scholman et al., 2017; Simner & Pickering, 2005; Scholman et al., 2020; Tskhovrebova et al., 2022a). Typically, in a continuation task, readers are asked to provide a logical and grammatical continuation to a prompt. These continuations are then categorized according to the factor of interest, in our case discourse relations. Expectations elicited in off-line tasks have been found to correlate with on-line measures, such as reading time and EEG responses (Kutas & Hillyard, 1984; De Varda et al., 2023; Scholman et al., 2017).

In Chapter 6, we investigate readers' **preferences for how discourse relations are expressed**. Here, we used a two-alternative forced choice paradigm rather than a numerical task, because the first has been argued to be better able to detect quantitative differences between conditions (Schütze & Sprouse, 2013). This task has been used before to examine preferences in discourse relation marking (Crible & Demberg, 2020). Numerical tasks, such as Likert scale ratings, are used in pretests for some experiments (e.g. in Chapters 5 and 6) to establish qualitative differences between conditions, in which we were not necessarily interested in whether there was an effect, but rather how large such an effect would be (cf. Scholman et al., 2017; Asr & Demberg, 2020).

2.4.2 Examining on-line processing

Off-line tasks reveal what the discourse representation of the reader looks like, but do not show how this representation comes about. On-line tasks provide insight into the time-course of the construction of this discourse representation. An often used methodology in on-line processing research are reading studies, measuring the time participants spend reading a specific word or region. In general, the assumption behind such methodologies is that higher processing difficulty will result in longer reading times, also referred to as the *effort-time link* (Wilcox et al., 2023). It is derived from a more specific hypothesis about eye fixations in reading (Just & Carpenter, 1980): the *eye-mind link*, which assumes that a reader's fixation on a word reflects that they are processing that word. Below, we discuss the two methodologies for obtaining reading times in the present dissertation, self-paced reading and eye-tracking.

In Chapters 4, 5 and 6, I use *self-paced reading task* to collect reading time data on a region of interest. Specifically, I adopt the cumulative moving window paradigm. In this paradigm, the entire text is presented on the screen, but all words are masked. The participant reads all words one by one: A button press reveals the next word (or couple of words, i.e. chunk) and masks the preceding one. The time between two button presses is taken as a measure of the time needed to process the current chunk. However, studies often find spill-over effects, meaning that effects that are hypothesized to occur on a region of interest show up in later regions (cf. Smith & Levy, 2013; Boyce et al., 2020). This suggests that the processing of a previous chunk also happens after it has been masked. Furthermore, self-paced reading is less natural than normal reading. For example, readers cannot reread earlier parts of a text (but see Paape et al., 2022, for an adaption that allows for rereading). Finally, self-paced reading provides a single reading time measure, which conflates various processes (e.g. the time needed for word recognition as well as the time that would otherwise have been spent rereading). Nevertheless, the ease of use and low cost of the self-paced reading paradigm continue making it a popular method for collecting reading time data, especially because it has been proven to reliably detect effects not only in lab-based settings, but also in crowd-sourcing (cf. Enochson & Culbertson, 2015).

A more ecologically valid alternative to the self-paced reading paradigm is *eye-tracking-while-reading*, which is used in Chapter 5. In this technique, eye gaze is recorded while the participant is reading the text (see e.g. Rayner, 1998; Clifton et al., 2007; Vasishth et al., 2013). These gaze patterns are divided into fixations, during which gaze is relatively stable on one point, and saccades, during which the gaze moves from one fixation to another. From this data, various reading measures can be extracted for a word or area of interest (AoI). Measures such as first fixation duration, first-pass reading time and skipping probability are believed to reflect early processes, such as word recognition, whereas total reading time, rereading time and regression probability have been argued to reflect later stages of processing like integrating the information in the context (Clifton et al., 2007, but see Vasishth et al., 2013; Von der Malsburg & Angele, 2017, for criticism on this distinction). As such, eye-tracking-while-reading provides more information about the reading process than self-paced reading (see Von der Malsburg & Angele, 2017, for the importance of applying corrections when analyzing multiple measures). The reading process during eye-tracking better reflects natural reading than in self-paced reading, since participants can skip or reread parts of the text. However, eye-tracking data is much more costly to collect than (crowd-sourced) self-paced reading data as it requires expensive

equipment, each participant needs to be tested individually, and there is a higher chance of data loss due to calibration issues or technical issues.

2.4.3 Converging evidence

There are various methods for investigating discourse processing, including corpus research, as well as off-line and on-line experimental methods. As discussed above, each method has its advantages and disadvantages: some methodologies are more natural, whereas others provide more control over what is actually measured. Importantly, each method provides a unique insight into a different aspect of discourse processing. Rather than selecting (results from) a single method, different methods should be seen as complementary.

Combining evidence from various sources or methods can facilitate theory building. Graesser et al. (1994) argue that theories should investigated by (a) generating testable predictions and comparing these predictions to empirical findings from (b) think alound protocols and (c) behavioral measures. They use this "three-pronged method" to compare various theories on inference making. Sanders & Evers-Vermeul (2019) draw on findings from corpus research, language acquisition and on-line and off-line processing studies to show that the cognitive notions of causality and subjectivity distinguish various discourse relations. Similarly, Scholman et al. (2022b) argue that evidence from different modalities (production, representation, processing and acquisition) is needed to establish whether a relational distinction that is descriptively adequate is also cognitively plausible. Such an approach also reduces the risk of drawing conclusions that are too strong based on false positives and encourage replicability across different methods.

In this dissertation, I therefore combine various methodologies to examine which sources of information influence the processing of discourse relations, focusing both on on-line processing (Chapters 4, 5 and 6) as well as off-line representations (Chapters 5, 6 and 7). This method of converging evidence is most explicit in Chapter 6, in which I combine corpus research with both off-line and on-line experimental tasks to examine the how statistical cooccurrence between clause structure and discourse relations affects the processing of `result` relations, finding different effects across the methodologies. Furthermore, in the next chapter, I investigate what determines whether and how a discourse relation signal influences discourse processing, drawing on findings from various sources.

2.4.4 Statistical analysis

An important part of empirical research is analyzing the data. There are many ways in which experimental data can be analyzed. Here, I introduce the general method for data analysis in this dissertation, specifying the choices in using this method.

Experimental data in psycholinguistics is often analysed with (generalized or linear) mixed-effects regression models (Baayen et al., 2008; Quené & Van den Bergh, 2008; Barr et al., 2013). These models allow for taking into account item-level and participant-level variation in the dependent variable (i.e. random intercepts) as well as by-participant and by-item variation in the effects of predictors (i.e. random slopes) simultaneously (i.e. mixed-effects). This is important, since participants vary considerably in factors that significantly influence the response variable, such as reading speed or lenience in acceptability ratings, and might each be affected differently by experimental manipulations. Likewise, there is significant variation in items in psycholinguistic experiments, for instance with respect to syntactic structure, length and content. While some of these factors, such as length and frequency, are usually accounted for with separate predictors, not all of the item-related variation can be captured in this way. Mixed-effects regression models capture the structure of the data, by considering this variation. In addition, mixed-effects models have been shown to reduce Type I errors (Quené & Van den Bergh, 2008; Barr et al., 2013) and increase power (Baayen et al., 2008; Barr et al., 2013), compared to traditional approaches such as separate by-participant and by-item repeated measure ANOVA's. This approach was therefore also taken in the analysis of empirical data in this study (see Chapters 4, 5, 6 and 7).

In all studies in this dissertation, I aimed for a maximal random effect structure (cf. Barr et al., 2013), reducing the model in case of non-convergence. More specifically, I removed intercept-slope correlations as well as those random effects that explained the least variance in a step-wise manner until convergence was reached. All analyses in this dissertation, unless specified otherwise, are conducted in R (R Core Team, 2022) with Rstudio (RStudio Team, 2020). For mixed-effects regression, we used the `lme4` package (Bates et al., 2015). Significance of the predictors was evaluated using `lmerTest` (Kuznetsova et al., 2017) and post-hoc analyses of interactions were conducted with `emmeans` (Lenth, 2024). Visualizations were made using `ggplot2` (Wickham, 2016), `effects` (Fox, 2003) and `xtable` (Dahl et al., 2019).

2.5 Processing discourse relations

The previous sections reviewed how discourse relations can be annotated and how the processing of discourse relations can be investigated experimentally. Now we turn to what findings from previous literature obtained using these approaches tell us about the processing of discourse relations. How do readers infer discourse relations? And how do different types of discourse relations infer processing?

2.5.1 Stages of discourse relation inferencing

To establish coherence, readers need to infer how the segments are related. This process can be divided into two sub-processes. On the one hand, readers need to know what the relation type between the two segments is (i.e. whether it is `result`, `condition`, etc.). For example, in (17) below, the reader has to understand that the two sentences are causally related. This has been referred to as *propositional integration* (Cozijn et al., 2011).

(17) Mary was late to work, because there was a traffic jam.

On the other hand, the reader needs to establish the validity of this relation (does segment A truly implicate segment B?), relating it to their prior knowledge about the world and updating that knowledge if necessary (Noordman et al., 1992; Van den Broek, 2010). With respect to (17), this requires activating general knowledge about traffic jams causing delays and adding this explanation for Mary being late to the mental representation. This process has been termed *world-knowledge inference* (Cozijn et al., 2011). In the presence of a connective, propositional integration occurs earlier than world-knowledge inference (Cozijn et al., 2011, see Section 3.5.2). In the absence of a discourse signal, however, world-knowledge inference has to precede propositional integration (but see below for an alternative explanation based on cognitive preferences). This suggests that world-knowledge plays an important role in discourse relation inference (Noordman & Vonk, 1992; Noordman et al., 2015). In Chapter 7, I will investigate the role of domain knowledge on discourse relation interpretation empirically, and examine various strategies that readers have at their disposal.

World-knowledge inferences are generally made on-line, although this depends on the presence of discourse signaling (discussed in Section 3.5.2), the availability of the necessary information to the reader, and the reader's goals. For example, studies investigating the amount of inferences needed to establish the relation show that events

that are directly causally related (e.g. *John's brother punched him. The next day, he was covered in bruises.*) are processed more easily than events that require an additional inference (e.g. *John's brother was angry with him. The next day, he was covered in bruises.*) (Keenan et al., 1984; Kuperberg et al., 2011). This causal relatedness already influences the earliest stages of lexical processing (Kuperberg et al., 2011). Note, however, that in this experiment, participants were asked to judge coherence (similar to Noordman et al., 1992). In fact, Noordman et al. (1992) only find evidence for the use of world-knowledge information in inferencing when the reader's task is to examine the text's coherence (Noordman et al., 1992). Furthermore, if the required world-knowledge inference is not (yet) part of the reader's knowledge, such inferences are generally not made on-line, even if the information is made available in the text (Noordman et al., 1992). If information is available through the reader's knowledge base, inferences will be made during reading (Noordman et al., 1992; Noordman & Vonk, 1992).

2.5.2 Discourse processing across relation types

Previous work has shown that some relation types are more difficult to process than others. Sanders & Noordman (2000) compared the processing of `list` relations with `problem-solution` relations in a self-paced reading task with Dutch native speakers and found that reading times for the latter were shorter (see also Mulder, 2008, Chapter 5). Xu et al. (2018) also found a processing advantage for causal relations with Chinese speakers: In their study, `concession` relations are processed more slowly than `consequence-cause` relations (see also Murray, 1997; Britton et al., 2024, for similar effects in English and Italian). This aligns with ERP research showing a similar effect: `concessive` relations elicit larger P600 effects than `causal` relations (Xu et al., 2015). Within the class of causal relations, subjective causal relations have been found to be read more slowly than objective causal relations in various languages (for English: Traxler et al., 1997, for Dutch: Canestrelli et al., 2013; Noordman & de Blijzer, 2000, for Chinese: Wei et al., 2021). Furthermore, the order of the cause and effect influence processing effort: Noordman & de Blijzer (2000) show that `result` relations are read faster than `reason` relations. This is similar to a more general processing advantage of temporally related events presented in chronological order compared to anti-chronological order (Münte et al., 1998; Politzer-Ahles et al., 2017; Scholman et al., 2022a). In Chapter 4, we examine how these differences in processing difficulty across relation types interact with the presence of relation marking.

How can these differences in the processing effort of discourse relations be explained? One factor in which discourse relations vary is their cognitive complexity (cf. Sanders, 2005; Spooren & Sanders, 2008; Evers-Vermeul & Sanders, 2009; Knoepke et al., 2017; Hoek et al., 2017). In terms of CCR (Sanders et al., 1992; Hoek et al., 2019a), negative relations, such as `contrast` and `concession`, are more complex than positive relations (e.g. `addition` and `result`) and anti-chronological relations are more complex than chronological ones (Münte et al., 1998; Evers-Vermeul & Sanders, 2009). Similarly, subjective relations have been argued to be more complex than objective relations, because it requires establishing a subject of consciousness (e.g. Canestrelli et al., 2013). The cognitive complexity of discourse relation has been shown to explain the order of acquisition of connectives (Evers-Vermeul & Sanders, 2009). It also partly explains differences in processing effort of the various relations discussed above.

2.5.3 Cognitive biases in discourse relation expectations

Causal relations form an exception to the finding that cognitively complex relations are processed slower. Causal relations are cognitively more complex than additive relations, as an implication relation presupposes a conjunction relation (Evers-Vermeul & Sanders, 2009; Sanders & Noordman, 2000). Nevertheless, there seems to be a processing advantage for causal relations. In addition, clauses that are in a causal relationship are remembered better than those that are not (Myers et al., 1987; Sanders & Noordman, 2000; Trabasso & Van Den Broek, 1985). This stronger representation of causal relations is surprising given that faster processing often leads to weaker representations (Zwaan & Rapp, 2006), which has been referred to as the paradox of causal complexity (Sanders, 2005): Causal relations lead to faster processing and better representation, despite a higher cognitively complexity. As a solution to this paradox, Sanders (2005) proposes the *causality-by-default* hypothesis, which states that readers assume that consecutive discourse segments are causally related, since this would be the most informative text representation. Indeed, readers consider clauses that have a large number of causal connections with other clauses, and are part of the main causal chain, to be most important (Trabasso & Van Den Broek, 1985). The *causality-by-default* hypothesis assumes that readers first try to relate two clauses causally, and only consider a less informative relation when no causal relation can be inferred or there are strong cues against such an interpretation. Since a causal relation is always the first type of relation that is inferred, it has a processing advantage compared to different types of relations. Thus, this theory explains why causal

relations are processed so fast despite their complexity, but also why a causal relation is often inferred in the absence of any lexical cues. In Chapter 4, we provide further evidence for the causality-by-default hypothesis, by showing that the presence of a connective facilitates reading less in `result` relations than in `concession` relations.

Another theory that has proposed general biases for discourse relations is the *continuity hypothesis* (Segal et al., 1991; Murray, 1997). It states that readers prefer to interpret information in a text as being temporally and causally continuous. Several corpus-based and experimental studies have provided evidence for these hypotheses. For example, continuous relations have been shown to be signaled less frequently by a connective (Asr & Demberg, 2012; Jin & de Marneffe, 2015), which is argued to be caused by the fact that these relations will be inferred regardless of coherence marking. In addition, these hypotheses align with differences across discourse relations in their processing effort described above: continuous and causal relations are processed faster than non-causal and non-continuous relations.

In addition to cognitive complexity and cognitive biases for continuity and causality, a final explanation for the processing differences between relations could be that relations differ in how predictable they are (cf. Mulder, 2008). In this case, the processing difficulty of a relation is hypothesized to be related to its unexpectedness, in line with an information-theoretic view of language processing (Levy, 2008; Demberg & Keller, 2008; Wilcox et al., 2023). Readers have been shown to make predictions at different levels of language representation (Heilbron et al., 2022; Kuperberg, 2016), even at the discourse level (Rohde & Horton, 2014). The cognitive biases discussed above could be seen as general expectations for specific relations. Comprehenders have indeed been shown to have strong expectations for causality: In continuation studies, participants provide causal continuations in the majority of the cases (Murray, 1997; Simner & Pickering, 2005; Mulder, 2008). However, if the processing difficulty of the discourse relation depends on its unexpectedness, this should not only hold across different relations (e.g. comparing `concession` with `reason`), but also within the same relation across different contexts. In Chapter 5, we aim to provide further evidence for the hypothesis that the processing effort of a discourse relation is related to its predictability, by investigating directly how the surprisal of a discourse relation in a specific contexts relates to reading time.

2.6 Conclusion

Readers continuously aim to establish coherence (Hobbs, 1979; Sanders et al., 1992), which they do by inferring how different segments of a text are related. These links are referred to as *discourse relations*. Discourse relations can be classified in many different ways (Mann & Thompson, 1988; Sanders et al., 1992; Webber et al., 2019). In this dissertation, I follow the labels provided in the PDTB framework, but will also refer to the dimensions proposed by CCR to distinguish cognitively plausible classes of relations.

Discourse annotation can provide insight into the distribution of discourse relation and their signaling, as well as into discourse relation inference and processing. However, determining the type of discourse relation that holds between segments can be challenging, and depends on the relation type, the presence and type of linguistic marking (see Chapter 3) and background knowledge (see Chapter 7). It is therefore important to determine the reliability of the annotation, using metrics such as Cohen's kappa or Gwet's AC_1. Reliable annotation is a prerequisite for drawing conclusions from the data.

Discourse relation inference requires integrating the two propositions and checking the relation with existing world knowledge. Such inferences are easier with some relations than with others: negative and non-causal relations have been show to be more difficult to process than positive causal relations. Cognitive preferences for continuity and causality have been argued to influence how discourse relations are processed and represented, an issue explored further in Chapter 4. However, it is unclear if general expectations for these types of relations also explain context-specific differences in discourse relation processing. This will be examined in Chapter 5. The next chapter describes how signals of discourse relations influence discourse processing.

Chapter 3

Signals of discourse relations

3.1 Introduction

Discourse relations can be signaled in various ways. Commonly, a two-way distinction is made in how a discourse relation is marked: the relation is either expressed with a connective or the relation is left implicit (see e.g. Asr & Demberg, 2012; Webber et al., 2019). In recent years, studies have shown that these so-called implicit relations often also contain cues for discourse relations (e.g. Das & Taboada, 2018a). I will define discourse signals (used interchangeably with discourse cues) as all those linguistic elements that provide information about the discourse relation. This definition will be discussed in more detail below. Discourse relation signals thus include connectives and cue phrases as well as other types of linguistic elements that can be a cue for the discourse relation.

Discourse relation signals come in many forms, ranging from syntactic structures to graphical information. This chapter discusses previous research on when and how discourse relations are signalled to provide an overview of the different types of discourse cues and the mechanisms influencing discourse relation signaling. Based on existing definitions and classifications of connective and non-connective cues (Schourup, 1999; Stede et al., 2018; Hoek et al., 2019b; Asr & Demberg, 2013), I derive features that characterize different types of discourse cues. This decomposition shows that signals of discourse relations form a continuum on these various aspects, going beyond a binary classification of what can or cannot constitute a discourse connective.

I then go on to show that these features correspond to the extent to which the discourse signal influences processing and representation. Many studies have investi-

gated the effects of connectives on processing and representation (e.g. Millis & Just, 1994; Cozijn et al., 2011; Cain & Nash, 2011), revealing distinct mechanisms triggered by connectives. There is much less research on the role of other discourse signals in discourse relation inference and processing. I review the literature on the effect of discourse signals on different levels of processing to compare how connective and non-connective cues influence discourse processing and representation.

In sum, this chapter not only provides a more defined decomposition of discourse relation signals into features, but also discusses how these features relate to empirical evidence on the role of discourse cues on representation and processing. Furthermore, it identifies gaps in the literature with respect to how discourse signals influence processing across languages, relations and the type of cue, as well as the role of predictability in discourse relation processing.[1]

3.2 Defining signals of discourse relations

There are various definitions of connectives and discourse markers (see e.g. Prasad et al., 2008; Schourup, 1999). Here, we will use the definition of a connective provided by Stede et al. (2018) as a starting point, which can be paraphrased as follows:

Connective A connective is a lexical item or phrase that cannot be inflected and conveys a two-place relation, whose arguments are abstract objects.

Stede et al.'s (2018) definition first of all describes what a connective *does*: it conveys a discourse relation. With respect to the discourse relation itself, restrictions are placed on their propositional content and their number of arguments. In this way, lexical items and phrases that do not mark a discourse relation, such as one-place discourse adverbs, like *probably* and *and* in coordinating noun phrases, are excluded. For a more elaborate discussion on what constitutes a discourse relation and its segments, see Chapter 2. The requirement that a discourse signal should provide information about a *discourse relation* applies to all signals of discourse relations, not only to connectives. I will refer to this as the **relation criterion**.

Secondly, in order to convey a discourse relation, the connective has to *provide information* about the type of discourse relation. Different connectives express dif-

[1]This chapter is based on (and in parts identical to) the following publication: **Marchal, M.**, Scholman, M.C.J., Sanders, T.J.M. & Demberg, V. (in prep.). Discourse relation signals in discourse processing and representation: Characteristics of the signal and the reader.

ferent ways in which the arguments are related. Connectives can even convey fine-grained differences between discourse relations. For example, Dutch has distinct connectives for expressing non-volitional (*doordat*), objective (*omdat*), and subjective (*want*) causality (Hoek & Scholman, 2023; Canestrelli et al., 2013). In fact, the existence of a linguistic signal can be considered as evidence for distinctions between relational categories (Knott & Dale, 1994; Scholman et al., 2022b).

On the other hand, there are signals for discourse relations (e.g *and*, negation) that can occur in many different relations. To illustrate, negation has been argued to be a cue for `chosen alternative` relations (Webber, 2013; Asr & Demberg, 2015). However, it also occurs in many other relations, without necessarily signaling that relation. This raises the question how to determine whether a cue actually serves as a signal or is merely an artefact of the way the message is formulated.

The information provided by a cue about a discourse relation can be quantified as its normalized point-wise mutual information (npmi, Bouma, 2009), which is an information-theoretic measure of association (Asr & Demberg, 2013). A negative *npmi* means that the element and the discourse relation co-occur less frequently than expected based on chance. A positive *npmi* reflects that the signal occurs more often with the discourse relation than expected and could be said to signal the discourse relation. Only if a linguistic element occurs more often with some discourse relation compared to others, it provides information about that discourse relation. If not, it cannot be considered a discourse relation signal. I will refer to this as the **information criterion**: A discourse signal should co-occur with the discourse relation beyond chance.

Unlike Stede et al. (2018), we remain agnostic with respect to the overt form of the discourse signal, since we are not only interested in connectives, but in all linguistic elements that can signal discourse relations. I will return to these characteristics of a connective in the next section. Here, I provide a definition of discourse signals, that includes both connective and non-connective cues for discourse relations:

Discourse relation signal A linguistic element that provides information about the discourse relation.

I will use *discourse (relation) signals*, and *(discourse relation) cues* interchangeably.

Note that this description is solely based on a statistical definition, which can be derived from corpus work. Indeed, corpus research provides insights on which signals co-occur with discourse relations. However, they do not reveal whether readers are sensitive to these cues as signaling discourse relations (i.e. whether these cues

actually *function* as a signal). Possibly, not all signals of discourse relation influence processing. In addition to corpus research, we therefore need experimental evidence to examine whether these cues influence how readers construct a mental representation of a text. We will review literature on the role of discourse signals on discourse representation and processing in Sections 3.5 and 3.5.2. However, to investigate which discourse signals influence processing, we first need to know what signals have typically been distinguished and how they can be characterized. This is discussed in the next section.

3.3 Connectives and beyond: Categorizing discourse signals

Traditionally, a distinction has been made between connectives and cue phrases (i.e. lexical signals of discourse relations, cf. Danlos et al., 2018), on the one hand, and other signals of discourse relations, which we will refer to here as non-connective signals, on the other. Below, we will describe these two classes in more detail, as well as several subcategories that have been distinguished, before discussing five features that characterize different types of connectives and other discourse signals. These features allow us to present the different types of signals on a continuum, rather than adopt a binary classification of what can or cannot constitute as a discourse connective.

3.3.1 Connectives

Connectives are prototypical signals for discourse relations (Knott & Dale, 1994; Prasad et al., 2008). As Stede et al. (2018) suggest, they can consist of a single word (e.g. *because*) or a phrase. These connective phrases can be a multi-word expression (e.g. *as a result of*), parallel connectives (e.g. *if ... then*), conjoined connectives (e.g. *if and when*) or modified connectives (e.g. *especially because*, *a week before*) (Prasad et al., 2008; Stede et al., 2018). Connectives exist in languages of different typologies (e.g. Dutch, Bourgonje et al., 2018), Bangla (Das et al., 2020), Nigerian Pidgin (Marchal et al., 2021b) and Chinese (Wan et al., 2024). For many languages, inventories of connectives have been constructed (Stede et al., 2018). However, there are cross-linguistic differences in the form of the connective (Zeyrek et al., 2020; Marchal et al., 2021b). In Turkish, for example, relational meaning can be conveyed

by suffixal subordinators in a similar way as connectives can in English (Zeyrek et al., 2020).

Traditionally, in the PDTB framework, connectives come from three different grammatical classes (Prasad et al., 2008): subordinating conjunctions, coordinating conjunctions and adverbs. Given the wide variation in how discourse relations are marked cross-linguistically, such a strict definition based on grammatical class is not feasible (Stede et al., 2018, see also Zeyrek et al., 2020; Marchal et al., 2021b) and was therefore changed to non-inflected lexical items. Still, phrases that mark the discourse relation lexically, but can themselves be modified, such as prepositional phrases (e.g *for that reason*) and verb phrases (e.g *this caused*) do not adhere to this criterion. These phrases are typically not included as connectives in corpora (e.g. Prasad et al., 2008) or connective lexicons (e.g. Stede et al., 2018) and have been classified as *secondary connectives* (Danlos et al., 2018; Stede et al., 2018) or *alternative lexicalizations* (Prasad et al., 2008). Secondary connectives come from many different grammatical classes, such as nouns, verbs and prepositional phrases (Danlos et al., 2018). In some cases, they contain context-specific information (Rysová & Rysová, 2014), as in (18) below, in which case they are referred to as *free connecting phrases* (Danlos et al., 2018). Contrary to primary and secondary connectives, free connecting phrases can only occur in specific contexts.[2]

(18) It hadn't stopped raining all day. **Due to the weather,** Mary did not cycle to work.

Connectives do not always exclusively signal a single relation: many connectives are ambiguous. For instance, *and* can be used in many different relations including `expansion` and `precedence` and `result` relations (Crible, 2020; Spooren, 1997). Similarly, *but* can signal both `contrast` as well as `concession` relations (Asr & Demberg, 2020). In addition, *while* occurs in both `concession` and `synchronous` relations (Webber et al., 2019). These connectives are sometimes also referred to as underspecified (Spooren, 1997), polysemous or poly-functional (Crible, 2020; Zufferey & Gygax, 2020b; Zufferey & Degand, 2024).[3]

[2]Rysová & Rysová (2014) refer to this context independence as the *universality principle*.

[3]Zufferey & Degand (2024) distinguish polysemy (a connective can encode different relational meanings) from polyfunctionality, which refers to fact that a lexical item can have multiple functions, one of which can be marking a relation.

3.3.2 Non-connective signals

Discourse relations can also be signaled by cues that are not part of the class of connectives. There have been several corpus efforts to identify such non-connective cues (e.g. Duque, 2014; Das & Taboada, 2018a; Crible, 2022). In addition, insights on non-connective signals for discourse relations come from computational studies aiming to improve automatic classification of implicit discourse relations (e.g. Sporleder & Lascarides, 2008; Pitler et al., 2009). To date, the largest effort in annotating non-connective cues is the RST Signalling corpus (Das & Taboada, 2018a). This is an annotation layer on the already existing RST Discourse Treebank (Carlson & Marcu, 2001), which consists of 176,000 words and 20,123 relations. Das & Taboada (2018a) show that discourse cues come in many different forms. They classify these signals according to their linguistic features. For example, their taxonomy includes lexical, semantic, morphological, syntactic and graphical cues. A similar taxonomy is provided by Crible (2022), who distinguishes signals operating at the semantic, syntactic, sentence or a proposition-adjacent level. To illustrate the variety of non-connective cues, I will shortly discuss discourse relation signals from these categories as well as some other signals below. In the next section, I will propose a different taxonomy of discourse signals based on features that are relevant to both connective and non-connective cues.

Lexical cues are words or phrases that signal the discourse relation, such as *contingent* in the example below.

(19) Mary offered to make John pizza, but that was contingent on what time he would be home. `condition`

They are similar to and partly overlap with secondary connectives and alternative lexicalizations discussed in the previous section. However, this class also comprises cues where the primary function is not to signal the discourse relation, as in (19) above, where the lexical item's contingent main function is its propositional one. I will return to this feature in the next section.

An example of a **syntactic** cue is parallelism, in which the syntactic structure of the two clauses is similar, illustrated in (20). It has been argued to be a cue for `contrast` relations (Das & Taboada, 2018a; Crible & Pickering, 2020).

(20) Mary always goes to Vietnamese restaurants. John usually eats at Italian places. `contrast`

Similarly, subject-auxiliary inversion is a cue for `condition` relations (Das & Taboada, 2018a; Webber et al., 2019). [4]

(21) Had it not been for the weather, Mary would be having salad. `condition`

Tense has been argued to be important in signalling temporal relations (Grisot & Blochowiak, 2021; Evers-Vermeul et al., 2017). It has been classified as a **morphological** (Das & Taboada, 2018a), as well as a sentence-level discourse signal (Crible, 2022). In their corpus study of English and French, Grisot & Blochowiak (2021) find that verb tense is a significant predictor of the type of temporal relations in both languages. For instance, relations with a pluperfect in the second clause are often interpreted as `succession` relations.

(22) Mary ate the salad. John had eaten the pizza.

Semantic cues, like antonyms or meronyms, can also signal discourse relations, as in the `contrast` relation in (23).

(23) Mary loved anything sweet. John preferred salty snacks. `contrast`

In addition, there are verbs that elicit strong expectations for certain relations. Implicit causality and consequentiality verbs elicit expectations for cause and consequence relations respectively (Pickering & Majid, 2007; Kehler et al., 2008), possibly because they carry an empty slot that needs to be filled (Bott & Solstad, 2021). An example is provided in (24).

(24) John praised Mary. She had made a delicious pizza. `cause`

An example of a **graphical cue** is punctuation, like colons as a cue for an `instantiation` relation in (25), taken from Dale (1991).

(25) Many of the policemen held additional jobs: thirteen of them doubled as cab drivers. `instantiation`

All of the cues discussed above are segment-internal: they occur in one of the arguments. However, discourse relations can also be marked segment-externally. For example, quantifiers as in (26), adapted from Scholman et al. (2020), signal upcoming `list` relations.

[4]This syntactic construction, as well as other syntactic and lexical-syntactic discourse signals, have been included as a new category, AltLexC, in the third version of the PDTB (Webber et al., 2019).

(26) The woman experienced several unfortunate events last night. She got wine thrown at her by her dining companion. On the way home, she sprained her ankle. `list`

Finally, much of the research on discourse processing, focuses on written text. However, there are numerous sources of information beyond the written domain. For example, pause duration, articulation rate and pitch correlate with discourse boundaries and structure (Den Ouden et al., 2009; Tyler, 2013). Causally related segments are read faster and with shorter pause durations (Den Ouden et al., 2009) and prosodic cues differentiate between subjective and objective causal relations in English (Couper-Kuhlen, 1996; Hu et al., 2022). Moreover, gestures have also been argued to contribute to signalling discourse relations (Laparle, 2022; Scholman & Laparle, accepted; Hinnell, 2019).

3.3.3 Features of discourse relation signals

As shown in the previous section, there is much variation in signals of discourse relation. How can this variation be captured? Here, I propose five features that describe various aspects in which discourse cues differ. The first two features are derived from the function of a connective, which is to convey a discourse relation (Stede et al., 2018). Two further features can be derived from Stede et al.'s (2018) description of the connective itself. The final feature is specific to non-connective cues and is derived from a classification of relation signals by Hoek et al. (2019b). In Section 3.5, I will discuss how these features correspond with the signal's effect on discourse representation and processing.

Functionality

A connective's primary function is to convey how segments are related. Conversely, connectives do not contribute to the truth-conditions of the separate propositions, referred to as the *non-truth conditionality* characteristic of connectives discussed by Schourup (1999). We will refer to this as the *functionality* feature: whether a signal's primary function is to convey the discourse relation.

Although the function of connectives and cue phrases is exclusively to signal the discourse relation, this is not the case for non-connective cues. Connectives and cue phrases can be removed or modified without changing the meaning of the individual propositions, but this is not the case for most non-connective cues. To illustrate, modifying the tense of the Example in (41), repeated here as (27), changes not only

the interpretation of the discourse relation, but also the meaning of the individual segments, as shown in (28). This is because tense first and foremost has a propositional meaning.

(27) Mary ate the salad. John had eaten the pizza.

(28) Mary is eating the salad. John will eat the pizza.

Compare this to (29) and (30) in which replacing the connective *after* with *because* does not affect the interpretation of the separate propositions.

(29) Mary ate the salad, after John had eaten the pizza.

(30) Mary ate the salad, because John had eaten the pizza.

This feature thus distinguishes the class of connectives and cue phrases from non-connective signals of discourse relations. Non-connective cues can be defined as linguistic elements, whose primary function is not to signal a discourse relation, but nevertheless provide information about the discourse relation.

Informativity

While some connectives signal a very specific relation, other signals refer to a broader class of relations or signal different types of relations. This illustrates a second feature in which discourse relation signals differ, which will be referred to as the *informativity* feature. It refers to the fact that the discourse signal can be more or less informative in determining which relation holds between the arguments. In information-theoretic terms, the informativity of connectives (and other cues) can be operationalized as the uncertainty about the upcoming discourse relation given the connective (i.e. the entropy). The information provided by the connective can be quantified as the reduction in uncertainty with compared to without the connective (see Asr & Demberg, 2013).[5] Some connectives signal a specific relation, but many connectives are ambiguous, varying in their informativity. Non-connective cues can occur in many or even all types of relations. In other words, non-connective cues are overall less informative than ambiguous connectives. This feature thus classifies various discourse relations on a continuum (see also Table 3.1 in Section 3.3.4).

[5]Contrary to *npmi*, which reflects the amount of information a signal provides about a specific relation, this measure indicates the amount of information a signal provides in general.

Immutability

Stede et al.'s (2018) definition of a connective states that a connective cannot be inflected. From this, we derive the *immutability* feature: whether the linguistic form of some discourse signals can be modified (cf. is grammaticalized, Danlos et al., 2018). As discussed above, this distinguishes primary from secondary connectives. In addition, variation in the immutability of connective is also reflected cross-linguistically. For example, the English primary connective *if* translates to *kdyby* in Czech, which can be inflected for person and number (see Danlos et al., 2018), and would therefore be classified as a secondary connective in Czech. This raises the question whether the distinction between primary and secondary connectives (or connectives and alternative lexicalizations) is truly a binary one or should be seen as a continuum (Danlos et al., 2018).

A further deviation of the immutability feature are free connecting phrases, since they are modified to fit a specific context. However, a part of the free connecting phrase is immutable, reflecting the connective. This does not hold for non-connective cues. Semantic cues like antonyms will be fully context-dependent. Rather than a binary distinction, immutability is thus a scale.

Lexicality

Whether discourse signals are lexical is captured in the *lexicality* feature. This feature reflects whether the signal is lexical or not. Like non-connective signals, secondary connectives and free connecting phrases come in many different forms. However, according to Danlos et al. (2018), secondary connectives and free connecting phrases all have a *core unit*: a lexical head that expresses the discourse relation. In this sense, they are similar to primary connectives. Some non-connective cues, such as negation and antonyms, are also lexical. Unlike secondary connectives, their form is context-dependent (cf. the immutability feature), but the meaning of the discourse relation is still provided by a lexical element. Other non-connective cues, however, do not have any lexical features. Examples are morphological and syntactic cues like tense and parallelism. Lexicality can thus distinguish within the class of non-connective cues.

Agreement

In the identification of non-connective signals, Das & Taboada (2018b) build on a list of signals that had been identified as potential signals of discourse relations in previous research, complementing it with signals they identified in the first part of

their annotation efforts. These signals are all "compatible with" the discourse relation they occur in (p. 756). Similarly, in her study on the interplay between non-connective cues and discourse cues, Crible (2020) annotates those linguistic items as signals "that encode congruent information that reinforces the meaning of the relation" (p. 16). In some cases, there is indeed overlap between the meaning of the linguistic element and (parts of) the discourse relation (cf. Hoek et al., 2019b). For example, *not* and the `contrast` relation it signals share the negative polarity meaning (Hoek et al., 2019b). We will refer to this as the *agreement* feature, which states that the signal's primary meaning is similar to the meaning of the discourse relation.[6] There are also cases where the meaning overlap between the linguistic element and the discourse relation is not clear and the linguistic element can nevertheless be considered to be a signal. For example, a parallel structure can signal `contrast` relations, but does not carry any contrastive meaning on its own. In these cases, the basis for the signalling status is that it simply often co-occurs with the discourse relation (Hoek et al., 2019b), i.e. its informativity.

3.3.4 A continuum of discourse relation signals

In a broad sense, discourse signals comprise of all linguistic elements, written, spoken or gestured, that provide information about the relation that holds between abstract segments, following the relation and information criterion. Above, we have derived a set of five features in which signals differ:

- *Functionality*: the signal's primary function is (not) to convey how segments are related
- *Informativity*: the signal is more (or less) informative in determining the type of discourse relation
- *Immutability*: the signal can (not) be modified
- *Lexicality*: the signal is (not) lexical
- *Agreement*: the signal's primary meaning is (not) similar to the meaning of the discourse relation

[6]Note that this can also comprise cases which Hoek et al. (2019b) would refer to as *division of labor*, in which the use of the connective is redundant, but there is still meaning overlap with the connective.

Table 3.1: Overview of discourse signal features and examples to illustrate the different types of cues. • indicates that the feature is (strongly) present. Gray bullets show that the feature is less strong or absent.

	connectives					**non-connective signals**				
	primary		secondary							
	because	*and*	*for this reason*	*this caused*	*due to the weather*	*is contingent on*	*not*	tense	parallelism	..
functionality	●	●	●	●	●	●	●	●	●	●
informativity	●	●	●	●	●	●	●	●	●	●
immutability	●	●	●	●	●	●	●	●	●	●
lexicality	●	●	●	●	●	●	●	●	●	●
agreement	-	-	-	-	-	●	●	●	●	●

Some of these features refer to binary distinctions: the signal is either lexically headed or not. Similarly, a signal's primary function is either to signal the relation or not (*functionality*), although a distinction can be made in whether the signal *also* conveys propositional meaning. Other features should rather be seen as a continuum, with signals varying in the extent to which they adhere to these features. This is illustrated in Table 3.1. For example, signals are more or less informative with respect to the exact discourse relation (*informativity*) and they differ in the extent their overt form is context-dependent (*immutability*). In addition, the signal's primary meaning can be more or less similar to the meaning of the discourse relation.

Together, the features can also be used to present a continuum, with relation-specific primary connectives on the one end, adhering strongly to most or all features, and non-lexical non-connective cues on the other. As can be seen in Table 3.1, *because* strongly adheres to all features, negation only a few and parallelism not at all. Still, all these linguistic elements co-occur with discourse relations in a non-random way and thus provide information about the upcoming discourse relation. Furthermore, this continuum mirrors the traditional classification of primary connectives, secondary connectives and non-connective cues.

These features cannot only be used to categorize discourse signals, but possibly also to describe their efficiency in signaling the discourse relation. To help the reader

to infer a discourse relation, a discourse signal should be both easily identifiable and informative. The functionality, lexicality and immutability features describe how easily a signal can be identified: Lexical items that have the same form regardless of context, and do not have another function, can be more easily recognized as a discourse signal. Once identified, the informativeness of the discourse cue is determined by its informativity: if the cue always signals the same relation, its meaning can easily be inferred. In addition, if the discourse relation overlaps with the signal's primary meaning (agreement), the meaning can be derived more easily. We will come back to this point in Section 3.6 after a review of previous literature on the role of discourse signals in processing and representation in Section 3.5.

3.4 Signaling discourse relations

Before discussing how discourse signals influence processing and representation, I will discuss their occurrence in natural language. There are many different ways to signal discourse relations, but to what extent are these discourse signals used in natural language? And how do different types of signals interact? A little over half of the relations in the PDTB 3.0 (Webber et al., 2019) are signaled with a primary connective (50.8%) or an alternative lexicalization (3.4%). Das & Taboada (2018b) only find connectives in 18.2% of the relations in their corpus. This considerable difference is likely due to differences in the discourse segmentation (cf. Section 2.2.1). Regardless, these numbers show that a large proportion of discourse relations is not marked by a connective. This does not mean that the relation is not signaled. In fact, in the RST Signalling Corpus, for only 7.3% of relations no signal could be identified. Das et al. (2018) find that the vast majority of relations (74.5%) are signaled exclusively by non-connective cues.

A relation can also be signaled by multiple cues. In fact, Das & Taboada (2018a) find that in 7.6% of relations, non-connective cues occurred together with primary connectives. In addition, they argue that non-connective cues can either be independent, in that they both separately signal the relation, or combined, in that the interpretation of one signal as a cue is dependent on another cue (Das & Taboada, 2018a). This is similar to Hoek et al.'s (2019b) distinction between *division of labor* and *agreement*, in their discussion on the interaction between connectives and non-connective cues. In some cases, non-connective cues can make another cue (e.g. a connective) redundant, as in (31), where the `chosen alternative` relation is signalled by *not*, even if *instead* would not have been there.

(31) Mary did not have pizza. Instead, she had a salad.

Alternatively, non-connective cues can complement another cue in that they convey certain aspects of the discourse relation, but do not function independently. An example is *never* in (32), which, like the connective *even though*, has a negative meaning. However, the negation alone does not signal the `concession` relation. It is therefore in agreement with, but not independent from the connective.

(32) Mary never has pizza. Even though she really loves it.

The wide variety of ways in which discourse relations can be marked raises the question what determines how a relation is marked. From a Gricean perspective (Grice, 1975), speakers are only as informative as necessary. With respect to discourse relation signaling, this suggests that a speaker will only use a connective if the relation cannot already be inferred based on cues whose primary function is to convey propositional meaning (Das & Taboada, 2018b). This also applies to choosing a specific vs. underspecified connective (Spooren, 1997).

From an information-theoretic view, the marking of discourse relations has been shown to follow the principles of the Uniform Information Density hypothesis. This hypothesis states that information tends to be distributed uniformly across the speech signal (Frank & Jaeger, 2008). Similar to the Gricean Maxim of Quantity, this suggests that only those discourse relations should be marked with a connective that would otherwise be unexpected (Asr & Demberg, 2012, 2015; Yung et al., 2017). The expectancy of the discourse relation is influenced by cognitive biases, as well as cues in the preceding context. Asr & Demberg (2012, see also Jin & de Marneffe, 2015) find that causal and continuous relations, which are expected based on cognitive biases for causality (Sanders, 2005) and continuity (Segal et al., 1991), are indeed less likely to be marked explicitly than other relations. Similar findings come from studies on translation, showing that expected relations are more likely to be left implicit when translated to a different language (Zufferey, 2016; Hoek et al., 2017). Furthermore, Asr & Demberg (2012) show that causal relations are marked by more ambiguous connectives, whereas `comparison` relations are generally marked with relation-specific connectives. This also provides evidence for the role of uniform information density in discourse relation signaling: relation-specific connectives have higher surprisal and thus occur mostly in more surprising relations (e.g. `contrast`). Ambiguous connectives, on the other hand, have a much lower surprisal and occur in more expected relations (i.e. relations with lower surprisal).

Apart from more general cognitive biases that influence the expectancy of the discourse relation, the effect of non-connective cues on the explicit marking of discourse relation has also been examined. Following the same line of reasoning as above, Asr & Demberg (2012) also hypothesize that implicit causality verbs, which have experimentally been shown to elicit a strong expectation for causal continuations (Kehler et al., 2008), should be less likely to be marked with a connective. However, they did not find that `reason` relations containing these verbs were indeed marked less often.

In a similar study, Asr & Demberg (2015) investigate the explicitness of `chosen alternative` relations, which can be marked with the connective *instead*. These relations have been found to co-occur with specific signals such as negation (see also Webber, 2013). Asr & Demberg (2015) show that negation is indeed predictive of `chosen alternative` relations (as well as some other relations) and that the relation is more likely to be explicit when this cue is present in the first argument. Contrary to Asr and Demberg's (2012) findings on implicit causality verbs, these findings are in line with the Uniform Information Density hypothesis.

The presence of non-connective cues also interacts with the specificity of the discourse signal. Based on the Uniform Information Density hypothesis, Crible (2020) predicts that non-connective cues will occur more frequently with more ambiguous connectives. The non-connective signals will reduce the surprisal of the relation, which is thus more likely to be marked with a more ambiguous connective.[7] Indeed, in her corpus study, the connective *and* was more often accompanied by other signals for the discourse relation than more specific connectives, such as *for example* or *whereas* (see Crible, 2022, for similar findings).

Much of the research on non-connective cues in discourse relation signaling focuses on its interaction with marking by a connective (e.g. Crible, 2020; Hoek et al., 2019b; Asr & Demberg, 2015). However, there is some evidence that the occurrence of non-connective cues in implicit relations (i.e. in the absence of a connective) is different from that in explicit relations. For example, Sporleder & Lascarides (2008) train a discourse relation classifier by removing connectives from explicit relations and having the model predict the discourse relation on these implicitated relations. However, they find that the features that can predict the discourse relations in originally explicit relations do not generalize well to originally implicit relations. This suggests that it is important to examine the role of non-connective cues in the absence of a connective.

[7]Note that Crible argues the other way around, such that relations that are marked by a more specific connective do not require other signals. However, since the non-connective cues have a different primary function, they are less likely to be omitted.

In Chapter 6, I present an experimental study on the processing of a non-connective cue in the absence of a connective.

To sum up, relations that are not signaled by a connective are highly common. However, even in the absence of a connective, there are usually other cues that signal the discourse relation. How a relation is marked seems to be constrained by information-theoretic principles. More specifically, discourse relations that are more expected, either due to cognitive biases or due to the presence of non-connective cues in the context, are less likely to be marked explicitly with a connective (Asr & Demberg, 2012, 2015). In addition, such relations are marked with more ambiguous connectives (Asr & Demberg, 2013; Crible & Pickering, 2020). A notable exception, however, are Asr and Demberg's (2012) findings that implicit causality, a strong indicator of causal relations, does not reduce the likelihood of relation marking. This raises the question whether surprisal of the discourse relation is the only factor that influences marking. We will return to this issue in Chapter 5.

3.5 Discourse signals in representation and processing

The previous section discussed how discourse signals are used in natural language. The main question addressed in this dissertation is how sensitive readers are to such signals, as evidenced by the effect of these signals on discourse representation and processing. Here, I review previous literature on the role of connectives and non-connective cues in discourse representation and processing separately. In the next section, I will show how these findings relate to the features of discourse signals discussed in Section 3.3.4 above.

3.5.1 Discourse representation

Connectives

Since discourse relations are a crucial part of building a coherent mental representation and connectives are clear lexical signals of these relations, connectives should facilitate text comprehension (cf. Kleijn et al., 2019). However, experimental findings on the role of connectives on subsequent representations are mixed (see Kleijn et al., 2019, for an overview), with some studies showing a beneficial effect of connectives (Millis & Just, 1994; Degand & Sanders, 2002; Sanders et al., 2007) and others find-

ing no or even negative effects (Freebody & Anderson, 1983; Sanders & Noordman, 2000; McNamara et al., 1996). Several factors seem to influence the effect of relation marking on comprehension (cf. Kleijn et al., 2019). Firstly, the effect depends on text characteristics. Coherence marking facilitates comprehension in difficult but not in easy texts (Linderholm et al., 2000; Kleijn et al., 2019). Secondly, the influence of the presence of connectives and other cohesive devices on comprehension is reader-dependent. The benefit of the connective seems to be most pronounced for readers with higher reading proficiency (Ozuru et al., 2009; O'Reilly & McNamara, 2007, but see Kleijn et al., 2019). This could in part be attributed to individual differences in connective comprehension (Crosson & Lesaux, 2013, see also Zufferey & Gygax, 2020a; Tskhovrebova et al., 2022a; Scholman et al., 2024a). In addition, readers with less prior knowledge show better comprehension after reading high-cohesion texts, but the reverse has been found for high-knowledge readers (McNamara et al., 1996; McNamara, 2001; O'Reilly & McNamara, 2007; Kamalski et al., 2008). Thirdly, the effect depends on the way representation is assessed. Connectives seem to have mostly a local effect (Van Silfhout et al., 2015; Kleijn et al., 2019). That is, they facilitate comprehension of the arguments connected by the discourse cue and the relation that holds between them, but less so for global text comprehension. Finally, the effect of the connective depends on the relation type. Kleijn et al. (2019) found that the explicit marking of contrastive relations was most beneficial. In additive relations, however, connectives decreased comprehension. For causal relations, the facilitating effect was only marginally significant. Possibly, readers' bias for causality (Sanders, 2005) reduces the need for a connective for inferring the relation.

Studies investigating the effects of discourse signals often do not only manipulate connectives, but also other elements, like referential chains and elaborative information (see e.g. McNamara et al., 1996; Freebody & Anderson, 1983). It is therefore not clear to what the extent the effects described above can be attributed to connectives or also to non-connective cues. Kleijn et al.'s (2019) findings suggest, however, that connectives seem to influence comprehension beyond other cues. The contrastive relations in their texts contained non-connective cues in order to ensure a `contrast` interpretation even in the absence of the connective. Still, contrastive connectives facilitated comprehension.

With respect to underspecified connectives, readers have been shown to interpret them probabilistically (Asr & Demberg, 2020). Asr & Demberg (2020) find that readers' interpretation of *but* and *although*, which can both mark `contrast` as well as `concession`, aligns with the proportion that these connectives signal these relations

in corpora. This effect was replicated in a comparison of sentence-initial and sentence-final *although*, which also have different relation type distributions. This shows that readers are sensitive to the co-occurrence between connectives and discourse relations. In Chapter 6, I examine whether this is also the case for non-connective signals.

Non-connective signals

Relation-specific as well as underspecified connectives thus both influence discourse relation interpretation (Asr & Demberg, 2020; Hoek et al., 2021c). What about discourse signals that have a different primary function? There is some evidence that non-connective cues also facilitate discourse relation inference. For example, Crible & Demberg (2020) investigated the effect of non-connective cues (resultative verbs for `result`, parallelism and antonyms for `contrast`) on interpretation and found that the presence of these cues facilitated inferring the correct relation, especially in `contrast` relations. This effect might have been driven by the lexical cues, since in their experiment on parallelism as a cue for `contrast` relations, Crible & Pickering (2020) find no facilitating effect of the presence of this cue (nor a connective) on verification accuracy or speed of disambiguating the relation. This does not mean that syntactic structure does not influence discourse relation inferences: in a series of rating and forced choice tasks, Rohde et al. (2017) show that the presence of the complementizer *that* in sentences like *The professor noted that the student teacher did not look confident and (that) the students were poorly behaved* reduces the likelihood of a `result` interpretation between the first and second embedded clause. In addition, eventuality type, state duration and event complexity has been shown to affect inferences about temporal order (Dery & Koenig, 2015; Marx et al., 2024). With respect to spoken language, prosody can guide readers' interpretation of subjective and objective causal relations (Hu et al., 2023). Thus, readers take into account non-connective information when inferring discourse relations.

Finally, discourse signals have been shown to elicit expectations about upcoming discourse relations (e.g. Kehler et al., 2008; Asr & Demberg, 2020). To illustrate, in an off-line continuation task, Scholman et al. (2017) find that *on the one hand* leads readers to provide more `contrast` relations, if expectations about the contrast introduced by *on the one hand* have not been fulfilled yet. These expectations are also elicited by non-connective cues. Scholman et al. (2020) show that more `list` continuations are provided after quantity expressions in the context, although this effect was modulated by participants' linguistic experience. Tskhovrebova et al. (2023) find that the effect of contextual list signals persists in the presence of additive, but not in the

presence of result connectives. Similarly, implicit causality verbs elicit expectations for upcoming causal relations (Kehler et al., 2008; Bott & Solstad, 2021). Moreover, in multi-modal communication, gestures also influence how readers continue the discourse structure (Scholman & Laparle, accepted).

3.5.2 Discourse processing

Connectives

Connectives and cue phrases have repeatedly been shown to influence discourse processing. They have also been referred to as 'processing instructions' (Van Silfhout et al., 2015), as they inform the reader on how to relate the information in different propositions. Research has shown that readers integrate the meaning of the connective quickly during on-line processing. For instance, if the connective is not appropriate for the discourse relation, processing is disrupted (Murray, 1997; Canestrelli et al., 2013; Xu et al., 2018; Wetzel et al., 2021). Moreover, connectives that mark relations that are more difficult to process (e.g. `subjective` vs. `objective` relations) slow down reading immediately (Canestrelli et al., 2013) and elicit stronger brain responses (Politzer-Ahles et al., 2017; Köhne-Fuetterer et al., 2021). Furthermore, connectives can also elicit expectations about upcoming material. Readers show anticipatory looks to referents that are plausible given the connective (Köhne-Fuetterer et al., 2021) and the processing of this referent is facilitated (Köhne-Fuetterer et al., 2021; Xiang & Kuperberg, 2015). Similarly, *on the other hand* facilitates the processing of *on the other hand* (Scholman et al., 2024b), updating these expectations when they are satisfied (Scholman et al., 2017). We will return to the role of prediction in processing discourse relations in Chapter 5.

Many studies have compared the processing of the second argument in the presence vs. the absence of a connective or cue phrase. For instance, Sanders & Noordman (2000) conducted a self-paced reading experiment, manipulating the presence of a cue phrase signaling a `list` or `problem-solution` relation. The subsequent sentence was read faster when it was preceded by a cue phrase signaling the relation, compared to when no cue phrase was present. Evidence that connectives facilitate processing has been found in readers of different ages, including eight- and ten-year-old children (Cain & Nash, 2011), teenagers (Van Silfhout et al., 2014, 2015) and adults (e.g. Millis & Just, 1994; Cozijn et al., 2011). In addition, this facilitating effect of the connective on reading the clause following it has been replicated in a variety of languages, including French (Grisot & Blochowiak, 2017; Blochowiak et al., 2022), English (Millis & Just,

1994) and Chinese (Chen et al., 2019; Xu et al., 2018). However, the size of the effect of discourse cues has been hypothesized to differ across languages (Blumenthal-Dramé, 2021). We will explore this further in Chapter 4.

The facilitating effect of the connective mainly occurs in the region directly following the connective. In a series of reading experiments, both Millis & Just (1994) as well as Cozijn et al. (2011) found that the presence of a connective sped up reading in sentence-initial regions, but slowed down reading sentence-finally. This might be due to a reactivation of the first argument at the end of the sentence: Probes from the first sentence are recognized faster in the presence of a connective, but only when presented sentence-finally (Millis & Just, 1994). Cozijn et al. (2011) attribute the initial speed-up to facilitated *propositional integration*: the connective helps the reader to establish the discourse relation between the two clauses. The sentence-final slow-down is argued to be due to a *world-knowledge inference* process. Cozijn et al. (2011) show that readers are faster to verify inferences about the validity of the relation in the presence of a connective. Note however, that a sentence-final slow-down in the presence of a connective has not been found consistently, with some studies finding no such slow-down (Van Silfhout et al., 2014, 2015) or even a sentence-final speed up (Zufferey & Gygax, 2016), possibly due to differences between relation types. We will return to this issue in Chapter 4, which presents a study investigating whether the facilitating effect of the connective is relation-dependent, examining both sentence-initial and sentence-final regions.

Evidence for the processes induced by the connective also comes from different reading measures in eye-tracking. Van Silfhout et al. (2015) find that connectives lead to more, but shorter regressions in the region directly following the connective. Similar to the processes of facilitated integration and stronger inference described above, this shows that connectives encourage readers to integrate the parts of the discourse relation (evidenced by more regressions to previous material), facilitating this process (evidenced by shorter regressions). In the region in which the relation can be inferred, regressions are less frequent in the condition with a connective (Van Silfhout et al., 2015, see also Cozijn et al., 2011; Van Silfhout et al., 2014), likely because at the moment where the relation needs to be induced, the information from the first argument is more active due to the connective. Facilitating effects of the connectives also show up in early measures. Van Silfhout et al. (2014) and Van Silfhout et al. (2015) find significant facilitation of the connective on first-pass reading time in sentence-initial region (see Cozijn et al., 2011, for forward reading time). Initial processing could be facilitated because the connective facilitates integration. An al-

ternative explanation for early facilitating effects of the connective is that they do not necessarily reflect facilitated integration, but rather a decrease of processing effort due to the connective making upcoming material more predictable. We explore such an expectation-driven account of discourse relation processing (cf. Kehler et al., 2008) in Chapter 5.

There is some evidence that ambiguous connectives influence the processing of upcoming material, showing that readers are aware of their possible meanings. Underspecified connectives facilitate the processing of relations that they are more likely to signal (Asr & Demberg, 2020). In addition, when the connective allows for another (non-primary) interpretation, it does not block expectations for such a relation (Mak & Sanders, 2013). Conversely, when such an ambiguous connective cannot signal the expected discourse relation, it immediately blocks such expectations (Koornneef & Sanders, 2013). Similarly, ambiguous connectives disrupt processing of the following discourse relation when used inappropriately. Wetzel et al. (2021) find that readers are slower when an ambiguous connective is used inappropriately compared to when it is appropriate. Note, however, that one of the connectives they included only has a different meaning when used in a different position and is thus polyfunctional rather than polysemous (cf. Zufferey & Degand, 2024).

Readers thus also rapidly integrate the meaning of underspecified connectives. However, the presence of an ambiguous connective seems to affect processing less than the presence of a relation-specific connective. Cain & Nash (2011) find that relations connected by *and* were read slower than those connected by a temporal, causal or adversative connective respectively. Similarly, subjective relations are read slower following the Mandarin underspecified connective *suoyi* than after the relation-specific connective *keijan* (Li et al., 2017). Blochowiak et al. (2022), however, find no difference in reading times of relations signaled by a relation-specific or underspecified connective. Furthermore, Zufferey & Gygax (2016) show that the presence of the polysemous French connective *en effet* facilitates the processing of confirmation relations compared to when no connective is present.

Non-connective signals

These mixed findings on ambiguous connectives raise the question whether non-connective cues, which are also not specific, can facilitate processing of the intended relation. There is some evidence that cues other than connectives influence discourse processing, although studies on the role of non-connective cues during on-line processing are limited. Schwab & Liu (2020) manipulated the presence of a lexical (e.g. *true*)

as well as a contextual cue (whether a contrast was introduced in the context), which were hypothesized to elicit expectations about an upcoming `concession` relation. The results of their self-paced reading experiment showed that these cues independently facilitated the processing of the connective *but* in German, although for English only an effect of the lexical cue was found. These findings suggest that non-connective cues can facilitate the processing of a connective, similarly to connective cues (cf. Scholman et al., 2024b).

A signal similar to the contextual cue in Schwab & Liu's study (2020) are implicit causality verbs, which have been shown to elicit expectations for upcoming causal relations in off-line tasks (Kehler et al., 2008; Bott & Solstad, 2021). Such expectations also affect on-line processing. Rohde & Horton (2014) show that readers are able to anticipate upcoming discourse structure based on implicit causality verbs. In addition, connectives that are incongruent with such an expectation (e.g *but*, *and*) immediately disrupt processing, whereas causal connectives do not (Koornneef & Sanders, 2013). Mak & Sanders (2013) find that implicit causality verbs also facilitate the reading of the second argument, as shown in shorter first-pass duration early in the sentence and shorter regression-path duration later in the sentence. Unlike with connectives, sentence-final processing was also facilitated, at least for regression-path duration (Mak & Sanders, 2013). Similar facilitating effects of implicit causality verbs on the processing of causal relations have been found for relative clauses (Hoek et al., 2021b). Moreover, when these expectations are satisfied by an explanation in a relative clause, subsequent causal material is processed slower (Hoek et al., 2021a). These findings suggest that implicit causality verbs elicit expectations, resulting in facilitated processing of upcoming material, and that readers update these expectations accordingly (cf. Scholman et al., 2017, for similar processes with *on the one hand*).

With respect to the effect of non-connective cues on the processing of the relation itself, findings are mixed. Crible (2021) found that the presence of negation facilitates the processing of `concession` relations. More specifically, in the absence of a negation signal, `result` relations are read faster than `concession` relations, in line with previous findings. However, this difference disappears in the presence of negation: `concession` relations were read equally fast as `result` relations when the first segment contained negation. In another study, Crible & Pickering (2020) investigate lexical-syntactic parallelism as a cue for contrastive relations. The facilitating effect of parallelism was stronger when the clauses were connected by the ambiguous connective *and* (i.e., when there was no other cue for `contrast`), than when the connective *but* was used. This suggests that parallelism also contributes as a cue for `contrast`

relations. However, this effect was only found in a task where participants were asked to disambiguate each sentence, suggesting that this type of information is only taken into account when the text is processed deeply. Finally, Grisot & Blochowiak (2017) examined the role of verb tense as a cue for temporal discourse relations in French. Based on corpus research, they hypothesized that verbs in simple past tense would elicit expectations for chronological discourse relations, similar to connectives like *then*, and thus lead to faster processing of these relations. In their self-paced reading study, the presence of this tense did not facilitate the processing of chronological relations, whereas the presence of a connective did. However, the simple past tense in French is less frequent and less preferred than the *Passé Composé* with which it was compared. This may have mitigated the effect of this cue on reading times.

In sum, there is abundant evidence that connectives facilitate the processing of upcoming material, including the discourse relation. However, they also trigger inference processing, resulting in sentence-final slow-downs. For ambiguous connectives and non-connective cues, the evidence for their effect on processing is inconsistent. Furthermore, note that in all studies on non-connective cues described above, the effect of the non-connective cues was investigated in the presence of a connective (except for Rohde & Horton, 2014). However, the majority of relations is signaled by non-connective cues exclusively (Das & Taboada, 2018a). It is still unclear how non-connective cues influence the processing of discourse relations in the absence of a connective. We will examine this in Chapter 6.

3.6 Features of discourse signals and their effect on discourse representation and processing

As argued in Section 3.3.4, discourse cues form a continuum on various dimensions, with connectives on one end of the scale and non-connective cues on the other. More specifically, discourse cues differ in whether signaling the discourse relation is their primary function (functionality) and if not, whether their primary function is semantically similar to the discourse relation they signal (agreement). Moreover, connective and non-connective cues vary in how context-dependent their overt form is (immutability) and whether they have a lexical head (lexicality). Finally, some discourse cues occur in only specific relations, whereas others can signal a variety of discourse relations (informativity). These features are hypothesized to also relate to the effect of discourse signals on representation and processing. The functionality, lexicality and immutability features would help the reader to identify the signal as

Table 3.2: Overview of example discourse signals with their features, as well as their effect on representation or processing. For the features, black bullets indicates that the feature is (strongly) present. Gray bullets show that the feature is less strong or absent. For the findings, black squares indicates that there is strong evidence for the effect of this type of discourse signal, with ■ indicating smaller effects or mixed findings and ■ representing that no effect has been found. -: not applicable

	because; nevertheless	*and; but; when*	*this caused; due to rain*	*several; true*	*not*; antonyms	implicit causality	tense; eventuality	parallelism
functionality	●	●	●	●	●	●	●	●
informativity	●	●	●	●	●	●	●	●
immutability	●	●	●	●	●	●	●	●
lexicality	●	●	●	●	●	●	●	●
agreement	-	-	-	●	●	●	●	●
representation	■	■	■	■	■	■	■	■
processing	■	■	■	■	■	■	■	■

a discourse relational cue. Subsequently, the signal's informativity and agreement features determine how informative the signal is and thus whether the reader will be successful in using it to identify the intended relation. To examine whether these features influence the representation and processing of discourse relations, I reviewed previous research on the effect of discourse signals above in the sections above. Here, I summarize these findings and relate them to the features. Table 3.2 provides an overview of the different types of discourse signals, their features and findings on whether the cue has been found to influence representation and processing.[8]

Both connective and non-connective cues facilitate the representation of discourse relations. Even when a signal's primary function is not to signal the discourse relation, its form is not grammaticalized or it can occur in different discourse relations,

[8]A more elaborate version, which also lists the evidence of the effects of the various cues, can be found in Appendix B.

the cue can still help to infer the relation (e.g antonyms, Crible & Demberg, 2020). However, the strength of the effect is modulated by these factors. For instance, discourse signals facilitate inference more when they are more specific, as annotation agreement on underspecified relations is lower than that on relations signaled by relation-specific connectives (Hoek et al., 2021c). In addition, expectations elicited by connectives, whose primary function it is to mark discourse relations, are stronger than by non-connective cues: Scholman et al. (2020) find only 35% `list` continuations in the presence of an expression of quantity. For `contrast` relations following *on the one hand*, this was 79%. Moreover, it seems that if the signal is not lexical, it requires meaning overlap between the signal and the discourse relation for it to affect representation: the presence of a parallel structure did not enhance verification of `contrast` relations (Crible & Demberg, 2020). All factors thus seem to be important features for predicting a signal's influence on discourse representation.

With respect to discourse processing, connectives and cue phrases have repeatedly been shown to facilitate the processing of discourse relations (e.g. Xiang & Kuperberg, 2015; Cozijn et al., 2011; Sanders & Noordman, 2000; Van Silfhout et al., 2015), using a variety of methods. There is some initial evidence that this is also the case for non-connective cues. However, these effects seem to be weaker. For example, Crible & Pickering (2020) only find an effect of parallelism on the processing of `result` and `contrast` discourse relations in a task where participants were repeatedly asked to disambiguate the relation. In addition, the effect of underspecified connectives is smaller than of relation-specific connectives (Cain & Nash, 2011).

Immutability does not seem to play a strong factor in the extent to which discourse signals facilitate processing. Similar effects of discourse signals have been found with immutable connectives (e.g *because*, Cozijn et al., 2011) and context-dependent cue phrases (Sanders & Noordman, 2000). With respect to agreement, it is also unclear whether this affects the influence of the signal on processing. Implicit causality verbs clearly affects processing (e.g. Rohde & Horton, 2014), whereas this effect is less clear for tense (Grisot & Blochowiak, 2017). Thus, functionality (connective vs. non-connective cues), informativity (specific vs. ambiguous connectives) and lexicality (negation vs. parallellism) seems to play a role in determining whether a discourse signal will facilitate processing, but this is less clear for immutability and agreement.

3.7 Conclusion

Connectives are prototypical cues for discourse relations, but there are many different ways in which discourse relations can be signaled. In fact, although nearly all relations are signaled in some way, less than half of the relations are marked by a connective. How a relation is signaled depends on the expectancy of the relation based on cognitive biases and the presence of cues in the context.

A broad definition of discourse signals is that they refer to all linguistic elements that provide some information about the discourse relation. Discourse cues can be distinguished based on their functionality, informativity, immutability, lexicality and agreement. These factors influence the extent to which they facilitate discourse representation and processing: readers seem to be more sensitive to signals that adhere strongly to these features than to those that do not. More specifically, all features were shown to relate to the signal's effect on discourse representation and functionality, informativity and lexicality seem to be most important in determining a cue's effect on processing.

Although facilitating effects of non-connective cues have been found, it is still an open question whether non-connective cues induce the same processes as connectives. Since their primary function is not to signal the discourse relation, they might not trigger the same inference processes as connectives do. Furthermore, most studies on non-connective cues also include connectives. It is unclear how non-connective cues influence processing in the absence of a connective.

Finally, when examining the facilitating effect of discourse signaling on processing and representation, several factors need to be considered. Firstly, the effects of discourse signaling are relation-dependent. The presence of a connective seems to facilitate comprehension and processing most for relations that are more difficult to process (Köhne-Fuetterer et al., 2021; Kleijn et al., 2019). Secondly, for on-line processing, the strength of the effect depends on the region of measurement. Both connective and non-connective cues in the context have been shown to facilitate the processing of the connective itself (cf. Scholman et al., 2017, 2024b; Schwab & Liu, 2020). With respect to the processing of the content of the relation, the effect is more pronounced in earlier regions, compared to later regions (Cozijn et al., 2011; Millis & Just, 1994; Van Silfhout et al., 2014, 2015). Thirdly, there is individual variation in the extent to which connective influences representation (e.g. Kleijn et al., 2019). This does not only apply to connectives, but also to non-connective cues (Scholman

et al., 2020; Hu et al., 2023). These factors need to be taken into account when investigating the role of connective and non-connective cues.

In the next chapters, I will present four experimental studies, investigating the effect of connectives on discourse processing (Chapters 4 and 5) as well as readers' sensitivity to non-lexical sources of information for the inference of discourse relations (Chapters 6 and 7).

Part II

Understanding connectives

Chapter 4

Across relations and languages Connective facilitation interacts with relation, but not language

As discussed in Chapter 3, connectives facilitate the processing of subsequent material. Here, we examine whether this effect is dependent on the causality of relation and language typology. Previous studies have revealed relation-dependent differences in connective processing (Murray, 1997; Xu et al., 2015; Köhne-Fuetterer et al., 2021). In addition, connectives have been hypothesized to facilitate more in some languages than in others due to typological differences between languages: Speakers of analytic languages (such as English) are assumed to rely more on contextual cues and therefore be less affected by the presence of a connective than speakers of synthetic languages (such as German), who are presumed to rely more on lexical information (Blumenthal-Dramé, 2021). In this chapter, I present two self-paced reading experiments to examine whether the presence of a connective has a stronger influence on reading in German compared to English and in non-causal compared to causal relations. We did not find any interaction between language and relation marking in either experiment. The effect of relation marking depended on relation type, such that the presence of a connective significantly facilitated reading `concession` relations, but not `result` relations. We discuss how these findings relate to earlier research on the time-course of connective processing (Cozijn et al., 2011) and the *causality-by-default hypothesis* (Sanders, 2005). We conclude that readers' sensitivity to the presence of

a connective thus depends on relation type, but we find no evidence that this also depends on language typology.[1]

[1]This chapter is based on (and in parts identical to) the following publication: **Marchal, M.**, Hewett, F., Scholman, M.C.J., Shahmohammadi, S., Stede, M. & Demberg, V. (submitted). The facilitating effect of the connective is dependent on the relation, but not on language. *Discourse Processes*.

4.1 Introduction

Connectives are important signals for discourse relations. As discussed in Chapter 3, they influence discourse processing and representation in various ways. Previous research has shown that they facilitate text comprehension (Kleijn et al., 2019; Van Silfhout et al., 2015) and elicit expectations about upcoming discourse relations (Kehler et al., 2008; Scholman et al., 2017; Asr & Demberg, 2020). Furthermore, connectives attenuate N400 effects for words that are predictable based on the connectives (Köhne-Fuetterer et al., 2021; Xiang & Kuperberg, 2015). There is long-standing evidence that connectives also facilitate reading. Evidence from both eye-tracking-while-reading (Cozijn et al., 2011; Van Silfhout et al., 2014, 2015; Xu et al., 2018; Chen et al., 2019) as well as self-paced reading studies (Millis & Just, 1994; Sanders & Noordman, 2000; Cozijn et al., 2011; Grisot & Blochowiak, 2017; Xu et al., 2018) shows that connectives lead to shorter reading times of the clause following the connective. This facilitating effect of the connective has been attributed to facilitated 'propositional integration': the connective helps the reader to establish the discourse relation between the two clauses (Cozijn et al., 2011).

Here, we test whether the facilitating effect of the connective on on-line processing depends on the relation type and on language, following Blumenthal-Dramé (2021). [2] Firstly, readers have been argued to have a preference for relating text segments causally (i.e. the *causality-by-default hypothesis*, Sanders, 2005). This leads to the hypothesis that the effect of the connective is smaller in `result` relations compared to other relations, since readers will infer `result` relations regardless. Secondly, with respect to language-dependent differences in connective processing, Blumenthal-Dramé (2021) hypothesizes that the effect of the connective is larger in German than in English. Due to typological differences between the two languages, readers are assumed to rely more on linguistic information in German than in English. Finally, the language-dependent effect of the connective might also interact with the relation type: speakers that rely more on contextual rather than linguistic cues might resort more to default strategies, such as a preference for causality. If so, the difference in the effect of the connective in `result` relations compared to other relations should be larger in English than in German. In other words, there should also be a three-

[2] This study has recently been retracted (The Editors of Discourse Processes, 2024), after we notified the author about an error in her data analysis. The studies reported here were conceptualized as replications and carried out before these issues were discovered.

way interaction between relation marking, language and relation. To conclude, our hypotheses are as follows:

H1 The facilitative effect of the connective on reading is smaller in `result` relations than in other relations (i.e. `contrast`, `concession`).

H2 The facilitative effect of the connective on reading is smaller in English than in German.

H3 The interaction between relation marking and relation is larger in English than in German.

Below, we first discuss previous work that has examined the role of the discourse relation type and language in the processing of discourse relations. We then present the results of two experiments which examine the effect of connectives across relations in English and German. Experiment 1 compares the effect of the connectives 'so' and 'but' and their German equivalents on the reading times of the relations `result` and `contrast`. In a similar vein, Experiment 2 compares the effect of the connectives 'therefore' and 'still' on the reading times of the relations `result` and `concession`. We do not find any interaction between language and relation marking in our experiments. We do however find that the effect of relation marking is larger for `concession` relations than for `result` relations. We end with a discussion on how these findings compare to earlier research on relation-dependent differences on the effect of the connective (e.g. Murray, 1997; Köhne-Fuetterer et al., 2021), the time-course of connective processing (cf. Cozijn et al., 2011; Van Silfhout et al., 2015) and the *causality-by-default hypothesis* (Sanders, 2005).

4.2 Background

4.2.1 Relation-dependent differences in connective processing

A first research goal of the present study is to examine whether the effect of the connective differs across different discourse relations. As discussed in Chapter 2, some relation types require more processing effort than others. For example, `problem-solution` relations have been found to be read faster than `list` relations (Sanders & Noordman, 2000; Mulder, 2008) and `consequence-cause` require less processing difficulty than `concession` relations (Xu et al., 2018, 2015). One explanation for the processing ease of `causal` relations is the *causality-by-default* hypothesis (Sanders,

2005). This hypothesis states that readers assume that consecutive discourse segments are causally related, since this would be the most informative text representation. Comprehenders have indeed been shown to have strong expectations for causality: In continuation studies, participants provide causal continuations in the majority of the cases (Murray, 1997; Simner & Pickering, 2005; Mulder, 2008). Findings from corpus research that `causal` relations are less likely to be marked with connectives (Asr & Demberg, 2012) have also been attributed to the *causality-by-default* hypothesis: Since readers will infer a `causal` relation between sentences even when this is unmarked, marking is less informative in these types of relations.

Further evidence for the *causality-by-default* hypothesis comes from studies on the role of the connective in discourse processing. For comprehension, Kleijn et al. (2019) show that texts with contrastive connectives are understood better than those without such connectives. This effect was only marginal for causal connectives. Several studies comparing relations marked with concessive and causal connectives have found that participants show lower comprehension on concessive than on causal items (Köhne-Fuetterer et al., 2021; Xu et al., 2015), suggesting that causal relations are easier to comprehend in general. In addition, concessive connectives elicit a larger P600 response than causal connectives. This indicates that they require a stronger update of the discourse representation (Köhne-Fuetterer et al., 2021).

Previous research has also shown a differential effect of the connective on reading across relation types. For example, Murray and colleagues find that the presence of appropriate `concession` connectives facilitate reading more than the presence of `causal` or `additive` connectives (as described in Murray, 1997). Similarly, inappropriate `concession` connectives disrupt processing more than inappropriate `additive` and `causal` connectives. Furthermore, a facilitating effect of the connective has been found consistently across different connectives in sentence-initial regions (e.g. Millis & Just, 1994; Cozijn et al., 2011; Van Silfhout et al., 2015), but in sentence-final regions, causal connectives seem to slow processing down (Millis & Just, 1994; Cozijn et al., 2011). For other relations, however, no such sentence-final slow-down has been found (Van Silfhout et al., 2014, 2015). One study even finds a facilitating effect of a non-causal connective at the end of the sentence (Zufferey & Gygax, 2016).

Following Blumenthal-Dramé (2021), we aim to test the *causality-by-default* hypothesis by investigating whether the effect of relation marking is smaller in `result` relations than in non-causal relations. The reasoning for this is as follows: if readers assume a causal relation between sentences by default, the presence (vs. absence) of a causal connective should not provide any benefit. Such a default assumption is not

made in `concession` or `contrast` relations, so the presence of a connective in these relations should facilitate reading.

4.2.2 Cross-linguistic differences in connective processing

A second research goal is to examine whether the facilitating effect of a connective is different across languages. Although the facilitating effect of connectives has been shown for a variety of languages (see above), there is some evidence that the extent to which readers rely on such lexical cues may differ between languages. Schwab & Liu (2020) find language-related differences in the facilitative effect of lexical and contextual cues, using a pair of flexible German and English discourse markers, namely '*zwar...aber*' and 'true/sure...but', as in (33).

(33) James likes to run [outdoors, $\emptyset$]$_{contextual}$. [True, $\emptyset$]$_{lexical}$ he has a treadmill in the living room, **but** he often jogs in parks.

In a self-paced reading study, they find that comprehenders could predict the German connective '*aber*' ('but') on the basis of the preceding lexical cue '*zwar*' ('sure/true'). Reading times on the connective '*aber*' were shorter when the connective was preceded by the lexical cue '*zwar*' compared to when it was not. They find a similar effect in English, but this effect was delayed compared to German (it was found on the spillover region). Schwab & Liu (2020) also included a condition with a contextual cue, which enabled readers to predict an upcoming `concessive` relation based on incoherence in the context. However, this contextual cue only facilitated the processing of the connective in German and not in English. This suggests that English readers rely less on contextual cues than German readers. However, the authors acknowledge that the difference between languages could also be attributed to differences in data acquisition: The German data was collected in-lab, whereas the English data was crowd-sourced.

Contrary to Schwab & Liu (2020), Blumenthal-Dramé (2021) hypothesizes that English readers rely more on contextual cues, whereas German readers make more use of lexical cues for coherence. Blumenthal-Dramé (2021) ascribes the extent to which German and English speakers rely on different sources of information (i.e. lexical or contextual) to typological differences between the two languages under investigation. One aspect that languages differ in is the extent to which meaning is encoded in the linguistic signal, also referred to as the analytic-synthetic continuum. Synthetic languages are characterized by heavy inflection, which encode distinctions such as tense, syntactic role or word class. Examples of synthetic languages are Russian,

and, to a lesser extent, also German. To illustrate, consider the example in (34) below:

(34) Du magst deine Nachbarin.
you.1SG.NOM like.2SG your.2SG.FEM.ACC neighbor.SG.FEM
'You like your neighbor.'

Both the pronouns and the verb indicate person as well as number. In addition, the pronouns indicate case and there is a special morpheme for gender on the noun. On the other hand, more analytic languages (e.g. English, Mandarin Chinese) typically lack inflectional morphology and have a low morpheme-to-word ratio. In the strong sense, meaning in these languages often has to be inferred. Consider the English translation of (34) above. The pronouns, as well as the verb, do not specify the number. The reader would have to infer from the context whether *you* refers to one or more addressees. Similarly, *neighbor* is underspecified for gender, which would also need to be inferred. In German, this information is provided in the linguistic signal.

Given the fact that meaning is often encoded more explicitly in synthetic languages, speakers of such languages have to infer less than speakers of analytic languages. These typological differences with respect to the form of the language may influence processing strategies. Speakers of analytic languages may be more prone to infer meaning from contextual cues in general, whereas speakers of synthetic languages may rely more on the linguistic signal. With respect to connectives, this suggests that speakers of German, which is a synthetic language, are more affected by the presence of a connective. Speakers of analytic languages, such as English, however, are hypothesized to rely more on contextual cues, and therefore more easily infer the discourse relation when a connective is absent (Blumenthal-Dramé, 2021).

Indeed, evidence from corpus research shows that relations are signaled more often in German than in English. In parallel translated data, more connectives are added in English to German translation, and more connectives are removed in German to English translation (Becher, 2011; Yung et al., 2023). Whilst some reasons for this are specific to the act of translation, there are also cross-linguistic differences which could account for the higher level of connectives in German. In addition to grammatical differences, according to Becher (2011), German speakers tend to use more connectives than English speakers and German texts generally contain more connectives to comply with these communicative norms. The ubiquitous presence of connectives in German may lead German speakers to rely on them more in processing than English speakers.

4.3 The present study

Here, we present two self-paced reading experiments examining the effect of the connective on subsequent processing. We test the *causality-by-default hypothesis* (Sanders, 2005) by investigating whether the effect of relation marking is smaller in `result` relations than in other relations (H1). Secondly, following Blumenthal-Dramé (2021), we aim to examine whether the facilitative effect of the connective on reading is indeed smaller in English, an analytic language, than in German, a synthetic language (H2). We also hypothesize that the relation-dependent effect of marking is larger in English than in German (H3).

4.4 Experiment 1: Comparing `result` to `contrast` relations

In this first experiment, we compare the influence of the presence of a connective in `result` and `contrast` relations. Like in Experiment 2, we test native speakers of English and German.[3]

4.4.1 Methodology

Participants

In total, 220 participants were recruited via Prolific. Half of them were native speakers of English who were based in the United States, and the other half were native speakers of German and were living in Germany. None of the participants reported any known language-related disorders. After excluding some participants from further data analysis (see below), a total of 201 participants remained: 102 English (mean age: 37; female: 49) and 99 German (mean age: 35; female: 39) native speakers.

Materials

The materials consisted of 20 items with `result` target relations and 16 items with `contrast` target relations in a 2×2×2 design: language (we created English and German versions of each item), connective presence ('marking', explicit vs. implicit),

[3]Neither Experiment 1 nor Experiment 2 was preregistered.

and relation type (`result` vs. `contrast`). The target relations were embedded in a communicative context. Example items for each condition can be found in Table 4.1.

For the `result` items, the explicit and implicit marking differed in the presence of the connective 'so' in English and '*deshalb*' in German. Since the presence of a `result` connective influences word order in German, the tense of the second argument was always present perfect to keep the word order of the target region (in bold) constant across conditions. The auxiliary verb was included in the pre-critical region, as the connective follows the auxiliary in German. Although the syntactic structure of the target region in the `result` relations is different across languages, this does not confound our findings, since we are interested in the interaction of language with relation marking, rather than a main effect of language. Note that within each language, the syntactic structure is the same across relation marking. For the non-causal relation, we chose to include the coordinating connectives 'but'/'*aber*', which would allow the syntactic structure to be the same across conditions as well as languages. The connective 'but' is used most frequently in `contrast` relations (Asr & Demberg, 2020).

We use a region-by-region self-paced reading paradigm, to ensure a single reading time for all critical words in a sentence. This target region, presented in bold in the examples above, is also the region in which the relation with the preceding sentence becomes clear. This region is followed by a spill-over region, consisting of the connective 'and' and an auxiliary verb phrase introducing the final clause. This was done to prevent any wrap-up effects on the target region triggered by punctuation in this region. Since no material can follow the verb in German, this had to be another clause. To make the repetition of 'and' more natural, the target sentence was embedded in a conversational context. Some of the materials were adapted from Blumenthal-Dramé's items (Blumenthal-Dramé, 2021), but some were newly constructed.

Procedure

The experiment was hosted on PCIbex (Schwarz & Zehr, 2021). The items were distributed across different lists, such that every participant saw each item only once. The experimental items were interspersed with 31 fillers from an unrelated experiment, resulting in 67 items per list. Almost a quarter of the items was followed by a comprehension question (n=16). The study took between 15-20 minutes and participants received £3.

Table 4.1: Example stimuli. The explicit condition contained the connective (in italics), the implicit did not. The critical region is highlighted in bold, the spillover is underlined. Chunk breaks are indicated with |.

Language	`result`
English	Finn had a date | with his girlfriend Daisy | at a new restaurant | last week. | His brother asked him | whether | the food was good. | Finn answered, | "Not really. | Daisy | ate spoiled fish | and (*so*) | **she got stomach cramps** | <u>and had to</u> | throw up."
German	Finn hatte letzte Woche | ein Date | mit seiner Freundin Dagmar | in einem neuen Restaurant. | Sein Bruder fragte ihn, | ob | das Essen gut war. | Finn antwortete: | "Nicht wirklich. | Dagmar | aß verdorbenen Fisch | und hat (*deshalb*) | **Magenkrämpfe bekommen** | <u>und musste</u> | sich übergeben."
	`contrast`
English	Khalil | was telling Eduardo | that he and his girlfriend | disagree so often. | He asked | if Eduardo | always sees eye to eye | with his girlfriend. | Eduardo replied, | "I don't mind | a white lie | *{and, but}* Jodie | **is always honest** | <u>and was</u> | shocked | when she | heard me lie | the other day."
German	Khalil | erzählte Mattheo, | dass er und seine Freundin | sich so oft streiten. | Er fragte, | ob Mattheo | immer auf Augenhöhe | mit seiner Freundin sei. | Mattheo antwortete: | "Ich habe nichts | gegen eine Notlüge | *{und, aber}* Judith | **ist immer ehrlich** | <u>und war</u> | schockiert, | als sie | mich letztens | lügen hörte."

Analysis procedure

Data from participants (n=18) who scored less than 70% on the comprehension questions as well as from one participant who reported that they noticed typos was removed. Upon checking the materials we found inconsistencies in two English items: in one of the `result` items there was an inconsistency in the name and in one of the `contrast` items a typo occurred. These items were removed from further data analysis. Furthermore, we removed data from items (n=13) on which the participant spent more than a minute, as this indicates that they may have taken a break. We also removed reading times lower than 100 ms or above 2500 ms as well as log-transformed reading time values 2.5 SD away from the participant's mean.

Since some of the covariates are confounded with the factors of interest, we residualize the effect of these covariates by estimating these effects on the filler items. A baseline model with log-transformed reading times of the filler items (excluding the first and last region of each item) was fit with trial order, region position, region length and their interaction with language as well as by-subject random intercepts and slopes.[4] The estimates of this model were used to predict reading times for the experimental items. The difference between the observed and predicted reading times (henceforth diffRT) on the regions of interest was then regressed on the factors of interest. Binary predictor variables were deviation coded (German, implicit, `contrast`: -1). Continuous predictors were centered and scaled. We aimed for a maximal random effect structure, but removed intercept-slope correlations as well as those random effects that explained the least variance in a step-wise manner until convergence was reached.

4.4.2 Results

The output of the model can be found in Table 4.2 below and the fitted reading times (diffRT) in each condition are plotted in Figures 4.1 and 4.2. The results show that `contrast` relations were read slower than `result` relations in both regions. Note, however, that there is also a significant interaction between relation and language in the target region, suggesting that the effect of relation is larger in German than in English in this region. This effect could be due to the differences in clause structure: the target region in the German `result` relations was preceded by an auxiliary verb, which might have facilitated processing of the target region. Such an auxiliary verb

[4] log(rt) ∼ trial*language + position*language + length*language + (1 + trial + length || subject)

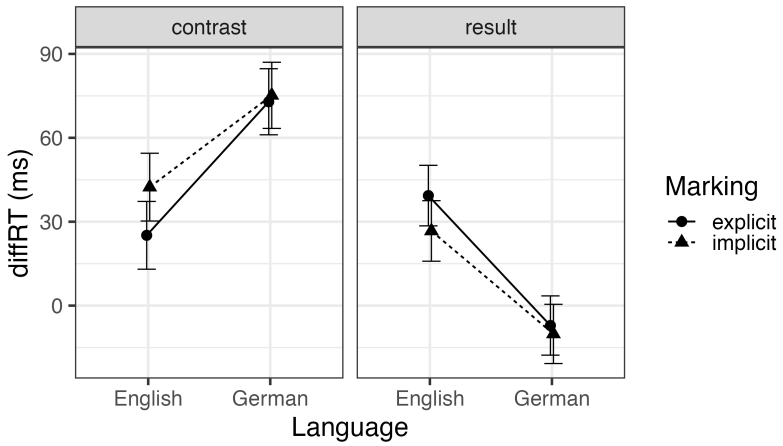

Figure 4.1: Fitted reading times (and standard error) of the target region per condition, relation and language in Experiment 1.

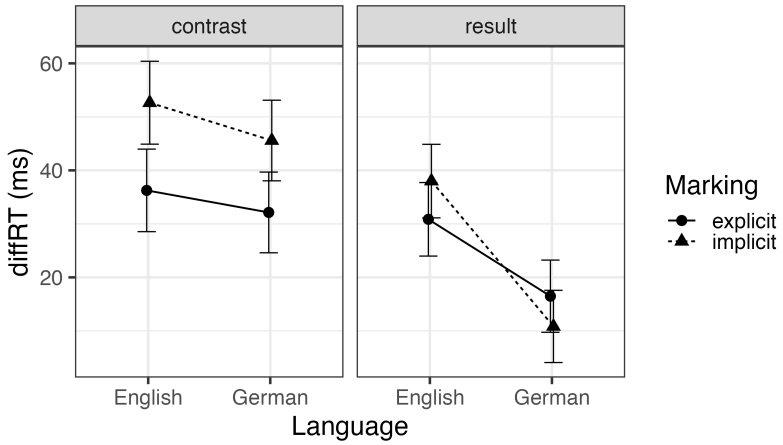

Figure 4.2: Fitted reading times (and standard error) of the spill-over region per condition, relation and language in Experiment 1.

Table 4.2: Model output of Experiment 1 for the target and spill-over region.

	target				spill-over			
	β	SE	t	p-value	β	SE	t	p-value
(Intercept)	33.05	5.47	6.05	**<.001**	32.85	3.38	9.71	**<.001**
marking	-0.48	3.14	-0.15	.88	-3.92	2.15	-1.82	.08
relation	-20.85	5.73	-3.64	**<.001**	-8.81	3.44	-2.56	**.01**
language	0.34	3.86	0.09	.93	6.59	2.56	2.58	**.01**
marking:relation	4.40	3.15	1.40	.17	3.54	2.15	1.64	.11
marking:language	-0.66	2.62	-0.25	.80	-1.97	1.77	-1.11	.27
relation:language	20.48	4.23	4.85	**<.001**	3.79	2.63	1.44	.16
mark:rel:lang	3.07	2.62	1.17	.25	-1.23	1.77	-0.69	.49

Model formula (target): `diffRT~mark*rel*lang+(1+mark*lang||item)+(0+rel|subj)`

Model formula (spill-over): `diffRT~mark*rel*lang+(1+mark+lang||item)+(0+rel|subj)`

was not present for the German `contrast` relations, nor in either condition of the English items. A post-hoc analysis reveals that the simple effect of relation is significant in German (β = 82.66, p <.001), but not in English (β = 0.73, p = .96).

Interestingly, we find no effect of relation marking on reading times in either the target or the spill-over region. The hypothesized interaction between relation marking and language is also not significant in either of the regions, nor is the interaction between relation and marking, or their three-way interaction with language.

One possible explanation for the lack of a main effect of relation marking is that the effect decreases throughout the experiment, because readers become familiar with the manipulation (cf. Fine et al., 2013; Hoek et al., 2021b). We therefore conducted a post-hoc analysis to explore whether trial number interacted with the effect of condition.[5] This two-way interaction was indeed significant (β: -5.17, SE = 2.35, t = -2.20, p = .03), such that readers sped up more throughout the study in the explicit items than in the implicit items. However, there was no effect of relation marking in either half of the experiment.[6]

[5]Model formula: `diffRT ~ marking*relation*language+trial*marking*relation + (1+marking*language||item)+(0 + relation||subj)`

[6]Nor the interaction between trial number and relation type nor its three-way interaction with relation marking was significant in either region.

4.4.3 Discussion

This study aimed to examine whether the facilitative effect of connectives on on-line processing differs across languages (English vs. German) and relation types (`result` vs. `contrast`). We did not find any evidence that the effect of relation marking differs per language. In fact, no effect of relation marking was found at all.

In line with previous research, `contrast` relations were read slower than `result` relations. However, this effect size was different across languages and may therefore have been confounded by the clause structure of the German `result` relations. Reading times were shortest in these items, which was the only condition in which the target region was preceded by an auxiliary verb. In addition, the effect of relation was only significant for German, but not for English.

The fact that we do not find an effect of relation marking raises the question if the lack of its interaction with language is mitigated by methodological limitations. We chose to examine an earlier region, since this is the disambiguating region and previous research has shown a more consistent effect of the connective in earlier regions (Cozijn et al., 2011; Van Silfhout et al., 2015). However, the effect shows up neither in the target region, nor in its spill-over, where it becomes apparent that a new clause has started.

In this experiment, the target clause was reported speech, preceded by a short context. Although reading such short texts could be argued to be more ecologically valid than reading single sentences, these longer excerpts did require participants to track more discourse entities. Possibly, this increased cognitive effort may have led readers to process the discourse relation more shallowly. In addition, this longer preceding context might have allowed readers to make better predictions about the upcoming clause. As a result, there might have been less room for facilitation driven by the connective. We note, however, that the effect of connectives and cue phrases has been established in longer contexts in previous studies (Sanders & Noordman, 2000; Van Silfhout et al., 2015; Cozijn et al., 2011).

Another possibility is that the connectives used in the present study were not salient enough. With the exception of German 'deshalb', all connectives were only two to four letters long (*so*, *and*, *but*, *aber*). Previous studies often use longer connectives or even cue phrases (Cozijn et al., 2011; Sanders & Noordman, 2000). The connectives tested here were chosen for their naturalness in a (reported) speech context.

In addition, the `contrast` relations we tested often had parallel structures. Crible & Pickering (2020) show that the effect of explicit relation marking ('and' vs. 'but') in

`contrast` relations is smaller when the two clauses have a parallel structure, providing experimental evidence for the observation that parallellism serves as a non-lexical signal for `contrast` relations, previously shown in corpus research (Das & Taboada, 2018b). Note, however, that this effect was only found when participants were required to overtly disambiguate the relation. It is unlikely that participants in our study processed the discourse relation so deeply.

We note that all of these factors might have influenced the effect of the connective, as well as its interaction with language and relation. We therefore conduct another experiment to test our hypotheses with a modified design, which is presented below.

4.5 Experiment 2: Comparing `result` with `concession` relations

Given that no effect of relation marking was found in Experiment 1, we test the same hypotheses in another experiment, with some methodological changes. In particular, we present participants with longer, more salient connectives, less context and a different non-causal relation. Here, we directly follow the design of Blumenthal-Dramé (2021).

4.5.1 Methodology

Participants

116 native speakers of English (mean age: 39, 56 male) and 144 native speakers of German (mean age: 36, 65 male) participated in the experiment. They were recruited from Prolific. Data from participants (n=6) who failed to answer less than 80% of the comprehension questions correctly were excluded from further analysis.

Materials

Stimuli were taken from Blumenthal-Dramé (2021) with minimal modifications.[7] As in the original study, the materials consist of 44 English items and 32 German items. The target relations consist of two coordinating clauses, connected by *and* (German: *und*). In each language, there are four different versions for each item, depending on

[7]We correct two typos and changed the wording of a few items (n=4) in German to make them clearer.

their discourse relation and the presence of a connective. The clauses form either a `result` or a `concession` relation, which is achieved by manipulating the first clause. Within each item, the second clause is the same across conditions. Secondly, each item is presented in an explicit (i.e. marked by a connective) and an implicit version. Example stimuli can be found in Table 4.3. In the explicit condition, `result` relations are marked by the connective 'therefore' and `concession` relations are expressed with 'still' in English. As per the original German stimuli, two different connectives are included per relation: *trotzdem*/*dennoch* (still) for the `concessive` and *deshalb*/*daher* (therefore) for the `result` relations. We add these connectives to all stimuli and therefore have two different sets of the 32 items for the explicit condition for each relation. Participants are shown equal numbers of stimuli with each of the different connectives, i.e. 50% of the explicit `concessive` sentences shown to a participant contain '*dennoch*', the other 50% contain '*trotzdem*'. No connective (except the coordinating conjunction 'and') was present in the implicit condition. The second clause of each item contained the target region, which was used for further analysis. Following Blumenthal-Dramé (2021), the target region consisted of a critical word – a lexical item at which the relation could be established – its spillover word and the final word of the sentence.

In addition to these experimental items, 256 filler sentences were created in English and 80 in German.[8] The fillers differ to the original study, but similar to the critical sentences, they have third person subjects with proper names. They do not contain causal connectives.

Procedure

A word-by-word moving window self-paced reading task was implemented using Ibex (Schwarz & Zehr, 2021). Participants were instructed to read in their natural pace with the aim of understanding the sentences fully. After a short practice phase, participants read a total of 96 sentences in German (16 experimental items) and 344 sentences in English (88 experimental items).[9] The order of these sentences was randomized for each participant. The German items were distributed across four

[8]The original study used 96 fillers in German. 16 of the German fillers in the current experiment were similar to the experimental items, except that the coordinating conjunction 'and' was replaced by a comma for a related experiment that is not reported here.

[9]Under a rapid expectation adaptation account (Fine et al., 2013), the smaller number of fillers in German would be expected to mitigate the effect. However, we would expect the effect of relation marking to be larger in German, thus the number of fillers bias against our hypothesis.

Table 4.3: Example stimuli. The explicit setting contained the connective written in *italics*, the implicit did not. The critical word is highlighted in bold. The stimuli are from Blumenthal-Dramé (2021).

Language	`result`
English	Luis won a prize and (*therefore*) **invested** in new equipment.
German	Anita war hochbegabt und übersprang (*daher/deshalb*) zwei **Klassen** auf der Grundschule. 'Anita was very gifted and (*therefore*) moved up two **years** at primary school.'
	`concession`
English	Christopher had no money and (*still*) **invested** in new equipment.
German	Anita hatte eine Lernschwäche und übersprang (*trotzdem/-dennoch*) zwei **Klassen** auf der Grundschule. 'Anita had a learning disability and (*still*) moved up two **years** at primary school.'

different lists such that each participant only saw each item once, with an equal number of items from each condition (relation, presence of connective). Following the original study, the English items were distributed across two different lists, where each list contained an equal number of items from each condition (relation, presence of connective), but contained the target clause twice (once in a `concession` and once in a `result` relation).

About a quarter of the sentences were followed by a comprehension question with two alternative answer options.[10] The German study took on average 20 minutes to complete and the English study around 45 minutes. Participants were compensated at least £9 per hour.

Analysis Procedure

We removed reading times below 100 ms and above 2000 ms, as well as log-transformed RTs that were more than 2.5 SD away from the participant's mean. As in Experiment

[10]Since the comprehension questions were not openly available, they differed to those used in the original experiment, but were formulated to fulfil the same criteria.

1, we fit a baseline model on the filler data to estimate the effects of the covariates and predict reading times for the experimental items.[11] We calculate the difference between the predicted and observed reading times (diffRT). We then summed these diffRTs for each word of the target region in each item to approximate region reading times as in our Experiment 1.[12] As in Experiment 1, all continuous variables were scaled and centered and binary predictors were sum-coded (German, implicit, `concession`: -1).

4.5.2 Results

Unlike in Experiment 1, we find a main effect of relation marking (see Table 4.4). As illustrated in Figure 4.3, reading times are shorter in the presence of a connective.[13] This effect is larger in `concession` than in `result` relations. A post-hoc analysis shows that the presence of a connective leads to shorter reading times in `concession` relations (β = 21.22, SE = 6.92, $p < .01$), but not in `result` relations (β = -3.17, SE = 6.85, $p = .64$).

We find no interaction between relation marking and language. In addition, the hypothesized three-way interaction is also not significant. Numerically, the direction of this interaction is even opposite to what was hypothesized: the interaction between relation and relation marking is *smaller* in German than in English.

The German stimuli included two different connectives for each relation; *'dennoch', 'trotzdem'* for `concessive` and *'daher', 'deshalb'* for `result`. In a follow-up analysis on the German subset of the data, we found that the two connectives did not significantly influence reading times for neither the `result` relations ($p = .79$) nor the `concession` relations ($p = .13$).

4.5.3 Discussion

Since we did not find evidence for any of the hypothesized effects, nor an effect of relation marking, in Experiment 1, we set out to test our hypotheses with a different design in Experiment 2. This experiment showed an interaction of relation marking

[11] We excluded the first, but not the last word of the filler items, since the experimental items also contain reading times of the item-final word. Model formula: `log(rt) ~ trial*language + wordpos*language + length*language + (1 + trial || subj)`

[12] A word-by-word analysis of the target region is presented in the Appendix C.

[13] A post-hoc analysis shows no significant interaction between the effects of interest and trial number.

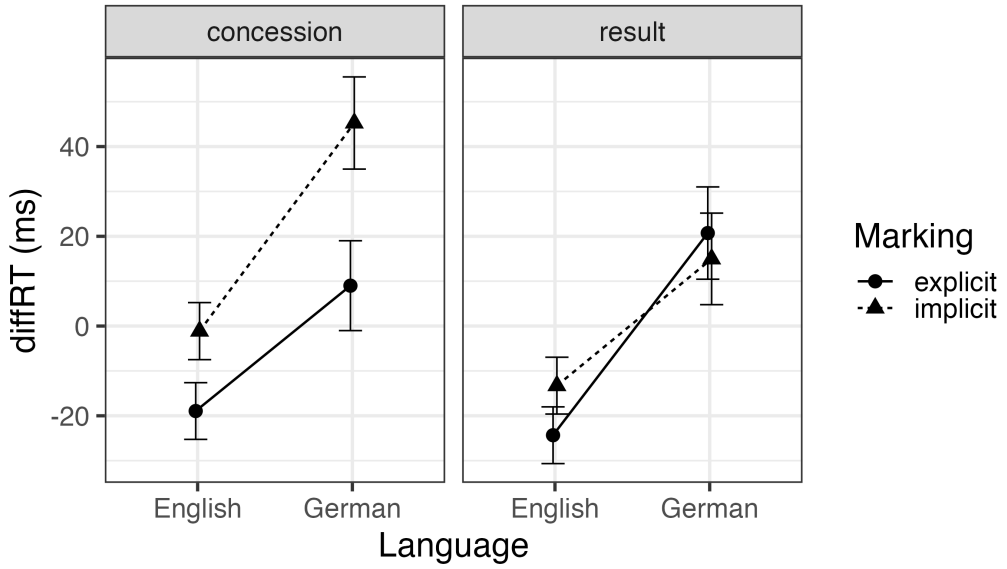

Figure 4.3: Fitted reading times of Experiment 2 per condition.

with relation, such that the effect of relation marking was larger in `concession` than in `result` relations. The connective facilitated reading in `concession` relations, but not in `result` relations.

The hypothesized interaction between relation marking and language was not found, nor did we find evidence for the hypothesis that the causality-by-default hypothesis holds in English, but not in German. Therefore, similar to Experiment 1, Experiment 2 also does not provide any evidence that the effect of relation marking, or its interaction with relation, differs across languages.

4.6 General discussion

This study set out to examine whether the facilitative effect of the connective is not only relation-dependent, but also language-dependent. More specifically, we aimed to examine whether connectives speed up reading times less in `result` relations compared to other relations, as assumed by the causality-by-default hypothesis (Sanders, 2005). In addition, we hypothesized that the effect of the connective is larger in German than in English, due to typological differences between the two languages (Blumenthal-Dramé, 2021). For this purpose, we presented two self-paced reading experiments. In Experiment 2, we indeed find that the effect of relation marking is larger in `concession` relations (H1). However, in neither of these experiments did we find a significant interaction between relation marking and language (H2). An

Table 4.4: Model output for Experiment 2.

	β	SE	t	p
(Intercept)	4.03	4.22	0.96	.34
marking	-7.42	2.52	-2.95	**<.01**
relation	-4.51	2.47	-1.82	.07
language	-18.45	4.22	-4.37	**<.001**
marking:relation	6.10	2.39	2.55	**.01**
marking:language	0.20	2.52	0.08	.94
relation:language	0.13	2.47	0.05	.96
marking:relation:language	-4.41	2.39	-1.84	.07

Model formula: `rt~mark*rel*lang+(1+rel||item)+(0+mark|subj)`

Table 4.5: Overview of results. Significance codes: $*p < .05$, $**p < .01$, $***p < .001$.

	Exp. 1		Exp. 2
	Target	Spill	Target
marking	n.s.	n.s.	**
relation	***	*	n.s.
language	n.s.	*	***
marking:relation	n.s.	n.s.	*
marking:language	n.s.	n.s.	n.s.
relation:language	***	n.s.	n.s.
marking:relation:language	n.s.	n.s.	n.s.

overview of the findings for each of the effects can be found in Table 4.6. Below, we discuss their theoretical implications as well as how the methodological differences between the studies might influence the findings.

4.6.1 The effects of relation marking and relation type

The main effect of interest is the influence of connectives on reading times. Connectives have been shown to facilitate processing for a variety of languages and relations, although this effect differs across relations and the time-course of the sentence (Millis & Just, 1994; Cozijn et al., 2011; Van Silfhout et al., 2015). In the present study, we find that the effect of relation marking is dependent on the relation, but do not find strong evidence that connectives facilitate reading overall (i.e. regardless of re-

lation). Data from Experiment 2 shows that the connective facilitates reading less in `result` relations than in `concession` relations. More specifically, we find that for `concession` relations, reading times are consistently shorter in the presence of a connective. For `result` relations, this is not found to be the case. No effect of the connective was found for `result` relations in Experiment 2.

One explanation for this finding comes from the *causality-by-default* hypothesis (Sanders, 2005), which states that readers assume a causal relation between consecutive discourse segments. A first prediction is that we should find a processing benefit for causal relations. In Experiment 1, `result` relations are indeed read faster than `contrast`. Note however, that this may have been confounded by differences in syntactic structure. In addition, we do not find a significant main effect of relation in Experiment 2. These findings of the effect of relation are thus not in favor of the *causality-by-default* hypothesis.

Secondly, if readers assume a causal relation by default, a causal connective, such as *therefore*, does not provide much additional information about the upcoming clause and will not facilitate reading. For `concession` relations, however, readers will need to update their assumption about the upcoming relation, and so the connective can provide more additional information about the content of the upcoming clause. In these cases, the causality-by-default hypothesis does assume a processing benefit for the connective. This is also in line with our findings: in Experiment 2, we found a facilitating effect of the connective for `concession`, but not for `result` relations.

Note, however, that in Experiment 2, the target region was sentence-final. Previous research has shown that connectives mainly facilitate reading in clause-initial, but not in clause-final regions (Cozijn et al., 2011; Van Silfhout et al., 2015). According to Cozijn et al. (2011), this is due to the connective triggering a world-knowledge inference at the end of the sentence, which requires more processing effort. Like the causality-by-default hypothesis, such an inference process could also explain why the connective does not facilitate processing for `result` relations in Experiment 2. Note, however, that we examined an earlier region in Experiment 1, but did not find an effect of relation marking in this region.

For `concession` relations, we do not see such a sentence-final slow-down induced by the connective. On the contrary, even in the sentence-final region, there is a beneficial effect of the connective. Thus, connectives *can* facilitate reading at the end of the sentence, although this effect is relation-dependent (cf. Zufferey & Gygax, 2016). With respect to world-knowledge inference, such a process is possibly not triggered sentence-finally for non-causal relations. For `concession` relations, inferences are

needed to establish the relation regardless of the presence of a connective. To illustrate, `concession` relations are also often referred to as `violated expectation` relations (Asr & Demberg, 2020). In order to establish the relation, the expectation that is subsequently violated needs to be inferred. This suggests that a world-knowledge inference will occur regardless of the presence of a `concession` connective. Still, the connective can facilitate processing through propositional integration. Possibly, the sentence-final facilitative effect of the connective in the case of the `concession` relations might be attributed to a spill-over of earlier processing difficulty of implicit relations, which was stronger for `concession` than for `result`.

4.6.2 Language-related differences in discourse processing

Experiment 2 shows that reading times are longer in German compared to English, after controlling for differences in word length. Note, however, that the items are different across languages and that we do not control for other factors that may differ between the two languages, such as frequency and surprisal. For example, the English stimuli mostly end on (highly frequent) temporal adverbs (e.g. *ran a marathon last month*, *lost the contest this summer*), whereas due to grammatical constraints the German stimuli cannot end on temporal adverbs. Instead, the German stimuli often contain additional (new) information in the spill-over region (e.g. *brought a spider from Sicily*, *finished the marathon without problems*), which does not only have higher surprisal, but also contributes to the interpretation of the relation. In Experiment 1, in which the items were the same across languages, reading times of the spill-over region were shorter in German than in English. However, in this experiment, syntactic structure differed across languages in the `result` condition. This illustrates that in cross-linguistic research, it is important to keep the material as well as data-collection procedure as constant as possible between languages.

Crucially, the effect of interest, namely the interaction between relation marking and language, does not reach significance in any of the analyses presented here. The hypothesis that the facilitative effect of the connective is different across languages therefore cannot be confirmed. Note, however, that the languages selected in Blumenthal-Dramé (2021) as well as in our replication attempts are from the same language family. In addition, they are not on extreme ends of the analytic-synthetic continuum. For example, Chinese has even less inflection and is thus more analytic than English. It is possible that differences in discourse-level processing emerge in languages that are more distinct. However, as discussed in the introduction, the facilitative effect of the connective has been replicated across language families in

many different languages, including French (Grisot & Blochowiak, 2017) and Chinese (Chen et al., 2019; Xu et al., 2018). Nevertheless, further research with more diverse languages is necessary to understand which aspects of discourse processing are language-specific or universal.

4.6.3 Limitations

Finally, we present two methodological considerations. First, the stimuli in the implicit condition all contain the word *and* (or *und* in German). Although *and* is an underspecified connective, it does provide some information about the upcoming relation. Importantly, it occurs more often within `result` than in `contrast` relations, and only occurs extremely rarely in `concession` relations (Webber et al., 2019; Das & Taboada, 2018b). Thus, even in the implicit condition, the `result` relation was already marked, albeit underspecified. This could also explain the lack of an effect in the `result` relations: the more specific `result` connective in the explicit condition provides only little additional information about the upcoming relation. With respect to the `concession` relations, however, no information is provided by *and* in the implicit condition. Thus, the `concessive` connective in the explicit condition signals the relation much more clearly than *and* in the implicit condition, compared to *therefore* vs. *and* for the `result` relations. This could also explain why there is a larger processing benefit of the relation-specific connective for `concession` compared to `result` relations.

Second, a methodological difference between the experiments in the present study is that Experiment 1 used chunked self-paced reading. The reason for this is that there is often not a single word that disambiguates the relation. For comparability, we therefore summed the reading times in Experiment 2. Still, it is possible that power is decreased in chunked compared to word-by-word reading. A further disadvantage of chunked self-paced reading is that this method is less time-sensitive. The word-by-word self-paced reading task allowed for further analyses on the time course of the effect, revealing for instance that the effect of relation in Experiment 2 reaches significance in the spill-over and sentence-final region, but not in the critical region, whereas its interaction with marking shows up in all regions but the final region. However, note that additional analyses also inflate the chances of a Type II error. It is therefore important that researchers have a priori hypotheses about where the effects show up, or conduct an analysis that takes this into account.

4.7 Conclusion

Connectives are an important source of information when establishing coherence (cf. Chapter 3), but how they influence discourse processing might depend on a variety of factors. In this chapter, we presented two experiments investigating two such factors: the causality of the discourse relation (corresponding to Research Goal 2 in Chapter 1) and the typology of the language (Research Goal 4). In both experiments, we fail to find evidence that the effect of connective on discourse processing is larger in German than in English. Thus, the present study provides no evidence that typological differences between languages, specifically the extent to which meaning is encoded in the linguistic signal, affects discourse processing. However, with respect to the causality of the relation, we do find differences in readers' sensitivity to the presence of a connective. Specifically, relation marking facilitated reading in `concession` relations, but we found no evidence for this in `result` relation. These findings are in line with the *causality-by-default* hypothesis. An alternative explanation would be that `result` relations are more expected than `concession` relations and that these differences in predictability explain the facilitating effect of the connective. This will be investigated in the next chapter.

Data availability

All items, data and analysis scripts of all experiments in this paper are available on OSF: `https://osf.io/6j8ew/?view_only=ce937096e97048cfaa7ec90aac137f0b`

Funding

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Grant 232722074 – SFB 1102 ("Information Density and Linguistic Encoding") and under Grant 317633480 - SFB 1287 ("Limits of Variability in Language"); German Ministry of Education and Research (BMBF) under Grant 01IS20058.

Dissemination

This chapter is based on the following journal article:

- **Marchal, M.**, Hewett, F., Scholman, M.C.J., Shahmohammadi, S., Stede, M. & Demberg, V. (submitted). The effect of the connective is dependent on the relation, but not on language. *Discourse Processes.*

In addition, (parts of) this research have been presented at the following conferences:

- **Marchal, M.**, Hewett, F., Scholman, M.C.J., Shahmohammadi, S., Stede, M. & Demberg, V. (2024). Connectives as processing instructions across relations and languages. *Talk at AMLAP, Edinburgh, Scotland, 5-7 September.*
- **Marchal, M.**, Hewett, F., Scholman, M.C.J., Shahmohammadi, S., Stede, M. & Demberg, V. (2024). Connectives as processing instructions across languages: A replication of Blumenthal-Dramé (2021). *Poster at Society of Text and Discourse, Chicago, Illinois, 24-27 July.*
- **Marchal, M.**, Scholman, M.C.J., & Demberg, V. (2023). Connectives as processing instructions across languages. *Talk at DGFS workshop Signalling discourse relations: Exploring (non-)connective cues', Cologne, Germany, 8-10 March.*

Chapter 5

Across levels of predictability Examining connective facilitation through relation and semantic prediction

Previous studies have argued that connectives help readers to integrate the arguments of the discourse relation (e.g. Cozijn et al., 2011, see also Chapter 3), as shown by the connective facilitating the processing of subsequent material (e.g. Sanders & Noordman, 2000; Van Silfhout et al., 2015, see also Chapter 4). Another explanation for this effect is that the presence of a connective allows readers to predict upcoming material more accurately. Indeed, comprehenders have been shown to make predictions at various linguistic levels (Kuperberg & Jaeger, 2016; Heilbron et al., 2022), but it is unclear whether discourse relation processing is also predictive. Here, we investigate whether the predictability of the discourse relation facilitates processing, teasing apart predictions about the relation type and the content of the relation. We obtain measures of relation type and content predictability from a continuation task and show that the presence of a connective increases the predictability of the relation type and, as a result, of the semantic content. In a self-paced reading and an eye-tracking experiment, we find that more predictable content is read faster. More predictable relation types also led to shorter reading times, but only because they facilitated predictions about the content. When the facilitating effect of content predictions was taken into account, more predictable relations resulted in longer

first-pass duration. The presence of a connective also facilitated reading, but we did not find evidence that this was due to the connective facilitating predictions about the upcoming relation type and content. Thus, we find no evidence that the role of connectives in discourse processing is dependent on the predictability of the relation.[1]

[1]The Pretest and Experiment 1 is based on (and in parts identical to) the following publication: **Marchal, M.,** Scholman, M.C.J., Sanders, T.J.M., & Demberg, V. (2024). What processing instructions do connectives provide? Modeling the facilitative effect of the connective. In *Proceedings of the 46th Annual Meeting of the Cognitive Science Society* (pp. 3435–3441). Together with Experiment 2, this work is prepared for publication in a journal: **Marchal, M.**, Scholman, M.C.J., Sanders, T.J.M. & Demberg, V. (in prep.). Predicting discourse relations: Understanding the facilitating effect of the connective.

5.1 Introduction

During language processing, comprehenders continuously predict upcoming material. For example, when hearing *The boy will eat*, listeners show anticipatory looks to objects that can plausibly follow this predicate (e.g. *the cake* and not *the ball*, Altmann & Kamide, 1999). This prediction also influences processing during reading. Predictable words (i.e. words with higher cloze probability) are read faster and are more likely to be skipped (Staub, 2015; Rayner et al., 2011). In line with an information-theoretic account of language processing, the predictability of a word is often operationalized as surprisal, the negative log probability of the word given the context. Indeed, there is substantial evidence that the relation between the probability of the upcoming word and its processing difficulty, as indexed by various measures, is logarithmic rather than linear (Smith & Levy, 2013; De Varda et al., 2023; Wilcox et al., 2023, 2024). Since surprisal refers to the extent to which the word *cannot* be predicted, we will refer to this as *unexpectedness* (cf. Heilbron et al., 2022).

Predictive language processing occurs at various levels (Kuperberg & Jaeger, 2016; Heilbron et al., 2022): reading times and brain responses have been found to be influenced separately by measures of phonemic, lexical, syntactic and semantic unexpectedness (Heilbron et al., 2022; Giulianelli et al., 2023). This processing is hierarchical, in that predictions about upcoming phonemes are constrained by information at other levels, such as syntactic and semantic information (Heilbron et al., 2022). Thus, readers predict phonemic, lexical, syntactic and semantic information and these various levels interact with each other.

Comprehenders have also been argued to make predictions at the level of discourse relations (Kehler et al., 2008; Rohde & Horton, 2014; Scholman et al., 2017, 2024b). However, the extent to which predictions about the type of discourse relation facilitates the processing of those relations during natural reading is still unknown. Earlier work on the processing of discourse relations has mostly focused on the role of bottom-up processes in establishing discourse relations (cf. Cozijn et al., 2011; Chen et al., 2019, see also Chapter 2, Section 2.5). The first goal of this study is to investigate an expectation-driven account of discourse relation processing during reading (cf. Kehler et al., 2008). More specifically, we examine whether discourse relation unexpectedness inhibits reading, while also controlling for semantic and local unexpectedness.

Secondly, the presence of a relation signal, such as a connective, has been shown to facilitate processing, especially in the region directly following it (Cozijn et al., 2011; Van Silfhout et al., 2014, 2015). This has been attributed to facilitated propositional

integration (Cozijn et al., 2011). However, this effect could also be attributed to the connective reducing the unexpectedness of the upcoming relation type and the semantic content of the second argument. The second goal of this chapter is therefore to provide more insight into the role of the connective in discourse processing, by examining whether connective facilitation is partly explained by reduced unexpectedness in the presence of a connective. Furthermore, we investigate whether there is an additional effect of the connective beyond effects of unexpectedness.

Below, we first review earlier findings on predictions at the discourse level as well as the effect of the connective on reading. Next, we discuss the operationalization of the unexpectedness of the upcoming discourse relation, semantic content and lexical-syntactic material in the current study and how it is influenced by the presence of a connective. We then present two reading experiments examining how unexpectedness at various linguistic levels as well as connective presence influences processing difficulty.

5.2 Background

5.2.1 Predicting discourse relations

In their expectation-driven account of discourse relations, Kehler et al. (2008) argue that readers are assumed to continually predict *which* discourse relation type (e.g. `result`, `contrast`) will follow. For example, when reading *Lisa was tired*, readers might make predictions about whether the text will next discuss *why* Lisa was tired (i.e. a `reason` relation) or what she is going to do about her tiredness (i.e. a `result` relation). Kehler et al. (2008) show that the interpretation of ambiguous pronouns depends on the discourse relation in which the pronoun occurs, in addition to the availability of the referent argued in earlier work. Since processing is incremental, readers must predict the upcoming discourse relation to arrive at an (initial) interpretation of the pronoun. Kehler et al. (2008) also suggest that expectations about the upcoming discourse relation will influence processing difficulty, but provide no empirical evidence for this.

Various studies have shown that material that is in line with the predicted discourse relation is processed faster. In line with Kehler's (2008) expectation-driven account, Mak & Sanders (2013) show that pronoun interpretation is indeed facilitated by expectations about causality. Pronouns referring to the first noun of the preceding sentence are processed faster when preceded by a verb that induced an expectation for

causality (as in 35), than in a context that does not elicit an expectation for causality (as in 36).

(35) The protester was fined by the policemen, when **he** broke the rules during the demonstration.

(36) The protester spoke with the policemen, when **he** broke the rules during the demonstration.

Such a facilitative effect on the pronoun is canceled when the expectation about causality is denied by the presence of a non-causal connective, such as *but* (Koornneef & Sanders, 2013).

Furthermore, empirical evidence that readers predict discourse relations comes from studies investigating the processing difficulty of connectives, which signal the upcoming discourse relation. For example, Scholman et al. (2024b) show that *on the other hand* is processed faster when it is preceded by *on the one hand*. *On the one hand* thus seem to elicit predictions of an upcoming `contrast` relation. This facilitating effect of *on the one hand* on *on the other hand* is modulated by whether the predicted `contrast` relation has already been satisfied (Scholman et al., 2017), suggesting that readers update their expectations about discourse relations (see also Hoek et al., 2021a). Similarly, Schwab & Liu (2020) show that the processing of the contrastive connective *but* is facilitated by lexical (*true/sure ... but*) and contextual cues (contrastive information) in the preceding context. Thus, expectations about discourse relations elicited by lexical and contextual cues facilitate the processing of connectives and sentence-initial pronouns that are in line with the predicted discourse relation.

Additional evidence that readers are indeed able to make predictions about upcoming discourse relations comes from a study by Rohde & Horton (2014). In a training phase, participants listened to sentence pairs with either a `cause` or `consequence` relation and saw a visual cue on either the left or right side of the screen, consistent with the discourse relation. Participants who had learned this mapping showed anticipatory looks to the location that the predicted relation mapped to based on an implicit causality or consequentiality verb in the prompt. This shows that readers can make predictions at the discourse-structural level and that these predictions are truly anticipatory. However, the paradigm is not very natural and might not generalize to natural reading. In addition, most of the studies discussed above make use of cues that strongly bias predictions for a specific discourse relation. Whether readers make predictions at such an abstract level during reading in less constraining contexts and how this influences the processing of the discourse relation itself is still unknown.

Processing different relation types

Different types of discourse relations have been shown to be more difficult to process than other relations. For example, `cause` relations are easier to process than `concession` relations (Xu et al., 2015, 2018), `problem-solution` relations compared to `list` relations (Sanders & Noordman, 2000) and `objective` relations compared to `subjective` relations (Traxler et al., 1997; Canestrelli et al., 2013; Wei et al., 2021). Similarly, connectives that mark relations that are more difficult to process have been shown to be more difficult to process themselves as well (Köhne-Fuetterer et al., 2021; Politzer-Ahles et al., 2017, see discussion in Section 3.5.2). In a bottom-up integration account of discourse relation processing, these differences have been attributed to differences in the cognitive complexity of the relation type. Following a Cognitive Approach to Coherence Relations (Sanders et al., 1992; Hoek et al., 2019a, see also Chapter 2), concessive relations are more complex than causal relations, because `concession` has a negative polarity.

However, these findings on processing differences between relations can also be explained by an expectation-driven account of discourse processing: the processing difficulty of a relation type is determined by its unexpectedness, with unexpected relation types yielding more processing cost (cf. Mulder, 2008). Indeed, the processing advantage of `reason` and `result` relations could be explained by their high predictability: In continuation studies, participants often provide continuations that are causally related to the prompt (Simner & Pickering, 2005; Murray, 1997; Mulder, 2008). Mulder (2008, Chapter 5) directly relates this to differences in the processing cost of various relation types. A continuation study showed that `problem-solution` relations are more expected than `cause-consequence` relations and `list` relations. In a subsequent reading study, `problem-solution` relations were read faster than the other two relations. This suggests that the predictability of discourse relation types could indeed explain differences in processing difficulty. However, previous studies have mostly focused on the differences in processing difficulty *between* different relation types. Here, we examine whether differences in unexpectedness can also explain differences in processing difficulty *within* the same relation type to avoid differences in cognitive complexity.

Predicting the content of the relation

During discourse relation processing, readers might not only make predictions about the upcoming relation type, but also of its content. Returning to the example of

Lisa being tired, readers might also predict whether the reason for Lisa's tiredness is that she did not get any sleep last night or that she had just finished a marathon. Further evidence for such predictions about the content of the relation is provided in Section 5.2.2 below. The unexpectedness of the relation type and of the semantic content have often been confounded in previous work (cf. Mulder, 2008). In fact, the unexpectedness of the content of the relation could also explain relation type differences in processing difficulty. The content of more unexpected relation types (e.g. `contrast`) might be more difficult to predict (i.e. there might be one likely consequence for a given event, but many different contrasting events). In order to find evidence for relation type prediction, the unexpectedness of the content should thus also be taken into account. We will refer to such predictions about the content of the relation as *semantic* predictions, as opposed to *relation* predictions which concern the type of relation that is predicted. Finally, how the second argument is formulated likely also influences the processing difficulty of the discourse relation. For example, less frequent (i.e. more unexpected words) have been shown to be more difficult to process than more frequent words (Scarborough et al., 1977). These predictions about the lexical-syntactic form of the content will be referred to as *local* predictions. We will discuss how we operationalize and distinguish different levels of unexpectedness in Section 5.4 below.

5.2.2 Connectives in discourse relation processing

Above, we argue that the predictability of the upcoming discourse relation type and content could facilitate processing. This expectation-driven account could also explain the facilitating effect of the connective found in previous research (Sanders & Noordman, 2000; Cozijn et al., 2011; Van Silfhout et al., 2015, see also Chapter 4). The clause-initial region, directly following the connective, is read faster in the presence of a connective, compared to when no connective is present Cozijn et al. (2011); Van Silfhout et al. (2015). No such effect, or even the opposite effect, has been found for clause-final regions, with longer reading times for sentences with a connective than with no connective (Cozijn et al., 2011; Van Silfhout et al., 2015). The clause-initial speed-up is attributed to the connective informing the reader how the clauses relate to each other (i.e. establishing the discourse relation, Cozijn et al., 2011). This has also been referred to as 'propositional integration'. Sentence-final slow down occurs due to the connective triggering a 'world-knowledge inference', in which the reader derives the relation and verifies its truth value using the reader's existing mental representation of the world.

In an expectation-driven account of discourse processing, the connective reduces the unexpectedness of the discourse relation, which in turn should facilitate processing. The processing gain provided by a connective can then be defined as the difference in this unexpectedness of the discourse relation with and without the connective. If a connective is not ambiguous (i.e. there is only a single relation that can (plausibly) follow the connective), the relation is fully predictable. However, in the absence of a connective, the relation type is more difficult to predict. Relation surprisal will thus usually be higher in the implicit condition than in the explicit condition. As a result, processing the relation will be more effortful and lead to longer reading times when no connective is present.

Connectives in different relation types

This expectation-driven account for the facilitating effect of the connective also predicts that the effect of a connective is larger when the relation type is unexpected in the absence of a connective - all other things being equal. In other words, the processing gain is higher when the connective signals an unexpected relation type. Indeed, previous research has shown that the benefit of a connective is smaller in some relation types than in others (Murray, 1997, see also Chapter 4). This has been attributed to cognitive preferences for interpreting text as being causal (Sanders, 2005) and continuous (Segal et al., 1991, i.e. an integration account), but these preferences could also be interpreted as general expectations for causal and continuous relations (cf. Murray, 1997; Asr & Demberg, 2012). As discussed in Section 3.4, these general expectations for causal and continuous relations have been shown to influence the presence of relation marking, in line with the Uniform Information Density hypothesis (Frank & Jaeger, 2008): the more unexpected a relation is, the more likely it is that the relation is marked by a connective (Asr & Demberg, 2012, see also Asr & Demberg, 2013, 2015). However, these experimental studies have not evaluated whether context-specific differences in discourse relation unexpectedness explain that the effect of connective presence is dependent on the relation type. Interestingly, Asr & Demberg (2012) do not find that the presence of an implicit causality verb, a contextual cue that has been shown to elicit expectations for causal relations (e.g. Kehler et al., 2008), influences the probability of relation marking.

Predicting the content of the relation

Furthermore, connectives also elicit expectations about the *content* of the upcoming segment. Köhne-Fuetterer et al. (2021) find that readers show anticipatory looks to

referents that are plausible given the connective. In addition, if the gender-marked adjective is not in line with this prediction, it elicits a larger N400 response (Köhne-Fuetterer et al., 2021). The processing of this predicted material is subsequently facilitated. EEG studies in both English and German have shown that connectives attenuate N400 effects for words that are predictable based on the connective (Köhne-Fuetterer et al., 2021; Xiang & Kuperberg, 2015). To illustrate, Xiang & Kuperberg (2015) had participants read sentence pairs as in (37) and (38).

(37) Meghan aced the test. She went home and **celebrated** wildly.

(38) Meghan failed the test. Even so, she went home and **celebrated** wildly.

They find that the N400 response to *celebrated* is smaller in the presence of the connective *even so*, than when there was no connective (as in Example 37). They argue that in the connective condition, readers could predict more accurately which event might follow.

In this scenario, the processing benefit provided by the connective is due to the upcoming material being semantically more predictable. The connective thus likely influences not only the predictability of the upcoming relation type, but also of the following content (cf. Jin & de Marneffe, 2015). This could also explain why Asr & Demberg (2012) find a higher proportion of connectives after *implicit causality* verbs, despite the fact that these verbs elicit expectations for causal relations. Although the *relation type* is predictable after an implicit causality verb, the specific *content* is not. Thus, both the unexpectedness of the relation as well as the semantic content seem to influence language processing.

Finally, the connective might influence expectations about how how the predicted discourse relation and semantic content is formulated at the lexical-syntactic level. This could in part also account for the facilitating effect of the connective, since there is likely a larger number of possible upcoming syntactic structures for a clause that does not start with a connective than for those that do. This factor will therefore also be taken into account.

5.3 The present study

In the above section, we saw that readers have been argued to make predictions at various levels of language processing (Kuperberg & Jaeger, 2016; Heilbron et al., 2022; Giulianelli et al., 2023), including discourse relations (Kehler et al., 2008). However,

empirical evidence for discourse relation prediction is limited, as previous studies focused on the role of integration in discourse relation processing. Furthermore, we aim to tease apart whether the processing difficulty of discourse relations can be attributed to the unexpectedness of the relation type or of the semantic content. More specifically, in this study, we investigate an expectation-driven account of discourse processing, by examining whether predictions about the relation type and the semantic content facilitates reading, while controlling for local measures of unexpectedness. Our first research question is as follows:

RQ1 Do lexical-syntactic, semantic and relation predictability influence processing during reading beyond other levels of predictability?

Furthermore, previous studies have shown that the presence of a connective leads to faster processing of the material directly following it (Cozijn et al., 2011; Van Silfhout et al., 2014, 2015). This benefit of the connective has been attributed to facilitated propositional integration (Cozijn et al., 2011), i.e. establishing which discourse relation holds between the two propositions. Here, we examine whether this effect can be attributed to predictions about the discourse relation type and its content. We hypothesize that the connective makes the upcoming material more predictable, which in turn facilitates processing. More specifically, the connective can elicit predictions about the local, semantic and/or discourse relation level. This predicts that (at least part of) the variance explained by the connective is explained through variation in local, semantic and relation predictability. Our second research question is thus:

RQ2 Can the processing benefit of the connective be explained by predictability at the relation, semantic or lexical-syntactic level?

Finally, it could be the case that the connective indeed enhances the predictability of upcoming material, but that this is not the only cause of connective facilitation and that other processes, such as propositional integration, also play a role. In other words, the connective might additionally influence processing difficulty directly, independently from facilitating effects of prediction. We therefore investigate a third research question:

RQ3 Is there an additional direct effect of the connective on processing beyond its effect through facilitated prediction?

5.4 Pretest: Operationalizing unexpectedness

In the present study, we will examine whether relation unexpectedness influences processing beyond other levels at which readers might make predictions. More specifically, we examine whether a `result` relation is read faster when this relation type is more expected given the context (i.e. has lower surprisal), independent of how expected the semantic and lexical-syntactic content of the `result` relation is. We do this by separately estimating relation, semantic and lexical-syntactic unexpectedness and examining how they influence reading times. In addition, we analyze how the presence of a connective influences each of level of unexpectedness and how this factor directly and indirectly predicts processing difficulty. In this section, we describe how we obtained different measures of unexpectedness. The relation between these factors and processing difficulty will be discussed in the next section.

We define relation unexpectedness as the negative log probability of the relation given the preceding context (i.e. relation surprisal):

$$Surp(rel) = -log2\ p(rel|context) \tag{5.1}$$

Secondly, we examine predictions about the *content* of the discourse relation, referred to as semantic predictions. To isolate semantic predictability, we consider semantic information value (Giulianelli et al., 2023). Semantic information value has been shown to predict processing difficulty similarly and complementarily to GPT surprisal (Giulianelli et al., 2023). It quantifies predictability of an utterance y given a context x as its distance d to plausible alternatives ($A_{context}$):

$$SIV(y|context) = d(y, A_{context}) \tag{5.2}$$

By considering how close the meaning of an utterance is to that of plausible alternatives, this measure captures that the content is semantically expected if other likely continuations carry a similar meaning.

Finally, we also consider the surprisal of lower-level lexical-syntactic material, since some syntactic structures or lexical wordings are more difficult to process than others. This is operationalized as empty context GPT2 surprisal. The preceding discourse context is not included to ensure that this measure is not conflated with higher levels of predictability.

5.4.1 Methodology

To estimate relation surprisal and semantic information value for the items in the reading study, we conducted a story continuation pretest.

Participants

One hundred and sixty native speakers of English (mean age: 39 years, range: 20-79 years; 87 female), recruited via Prolific, participated in the experiment.

Materials

The materials consisted of 24 items, which all had the following structure: (i) an introductory sentence, (ii) a sentence that was the first argument of the relation, (iii) a pronoun and (iv) a target region. This target region, which was not shown to the participants in the pretest, would later be presented in the reading study and disambiguated the discourse relation. There were four versions for every item, following a 2x2 design. All versions within an item ended with the same target region, but differed in (a) whether the sentence containing the target region started with a connective and thus the relation was signaled (i.e. explicit) or not (i.e. implicit), and (b) whether the target region was either highly predictable or less predictable depending on the context. The latter distinction was implemented to ensure variation in the semantic predictability of our items and was based on an earlier pretest (see Appendix D). This distinction is not relevant for subsequent analyses, since the goal of the current pretest is to obtain a *continuous* measure of relation and semantic predictability that is estimated for each condition and context separately. An example of the different contexts for an item is shown below:

(39) Angela used to live in a small flat in Atlanta.
   a. She didn't pay rent for months. {Therefore, she} / {She} ...
   b. She had over fifteen cats. {Therefore, she} / {She} ...

target: was evicted

Procedure

The items were distributed across 8 lists, each containing 12 experimental items (3 in each condition), as well as 18 fillers (6 implicit relations, 6 relations with "then" and 6 with "because"). Every list was completed by 20 participants, who were instructed to provide a logical continuation to the prompt.

Analysis

Data of a participant who repeated the prompt without providing a continuation was removed and replaced with that of a new participant. The continuations were analyzed for their relation and the event. Relation annotation was done by the first author, according to the PDTB3 guidelines.[2] All items in the explicit condition were coded as `result` relations. 10% of the data was double-annotated. Inter-annotator agreement was sufficient (k = .65 [.50, .79], AC_1 = .84 [.76, .91]). Relation surprisal was obtained by taking the log probability of the target relation per item per condition.

To estimate semantic predictability, we followed Giulianelli et al. (2023)'s approach. Continuations were first cleaned by removing typos and repeated pronouns. Subsequently, sentence embeddings were obtained for each continuation, as well as the target event, using Sentence-BERT (Reimers & Gurevych, 2019). For every item, we took the mean Euclidean distance between the continuations and the target region per condition.[3]

Relation surprisal and semantic information value both take into account the preceding context, but they do not account for the fact that the specific formulation of the expected relation or content also influences processing difficulty. For example, given a prediction that Angela will have to leave the flat, processing *was evicted* might be more or less surprising than *was kicked out*. For this reason, we added another measure of surprisal, using GPT-2 to calculate surprisal of the target region to capture item-related differences in local surprisal (LS). We excluded the preceding context when calculating LS to ensure that this LLM surprisal was not influenced by relation- and semantic-level predictions. We did include the connective as well as the pronoun in the preceding context, since some syntactic structures or lexical items might be more expected following a connective than others.

A prerequisite for the hypothesis that enhanced predictability explains connective facilitation is that relation and semantic unexpectedness are indeed lower in the presence of a connective. However, relation unexpectedness might also influence the unexpectedness of the semantic content: If the relation is unpredictable, the semantic content of that relation will be so too. As a result, the connective could also indirectly influence semantic unexpectedness by reducing relation surprisal.

[2]The features +BELIEF and +SPEECHACT were excluded.

[3]These parameters (mean vs. min; Euclidean vs. cosine) were selected, because they are the strongest predictors of reading times (see Table 6 in Giulianelli et al., 2023)

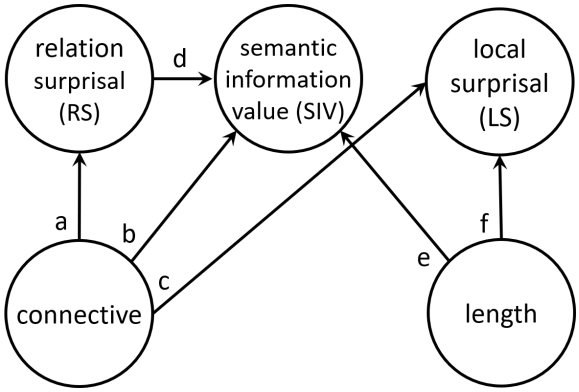

Figure 5.1: Structure of the piecewise structural equation model.

One way to model such indirect effects are structural equation models (SEM). This is a type of path model that allows for estimating hypothesized direct and indirect effects between predictor, mediator and outcome variables. The structure of such a model is often illustrated using arrows between the variables in the model (see Figure 5.1), with the arrows representing relationships between the variables. The variables with incoming arrows are referred to as *endogenous* variables (i.e. they are all variables that are predicted). Variables with both incoming and outgoing arrows are *mediator* variables. Direct effects (or single paths, e.g. connective → relation surprisal) are estimated as the standardized effect of a predictor on an endogenous variable. The indirect effect of a predictor through a mediator on an outcome variable (e.g. connective → relation surprisal → semantic information value) is then the product of these single paths. The total effect of a predictor on a response variable is the sum of all direct and indirect paths. Traditionally, these paths are estimated by estimating the covariances among all variables in the model simultaneously (i.e. global estimation). However, this requires a large amount of data in order to obtain sufficient power. Furthermore, traditional SEM approaches do not allow for complex variance structures like in multi-level regression models.

An alternative way of estimating the model is based on graph theory (Shipley, 2000). This approach is also referred to as *local estimation* or *piecewise SEM* (Shipley, 2000; Lefcheck, 2016). As the name suggests, in piecewise SEM, each path is estimated locally. That is, the model is made up of separate multiple regression models for each endogenous variable. To test whether the predetermined path model fits the data, tests of directed separation are conducted, which evaluate whether there are relationships in the data that are not specified in the model. These can be summarized in a single statistic of model fit: the *Fisher's C*. A non-significant C statistic, which is analogous to the χ^2 statistic with 2 degrees of freedom, reflects that the null hypothesis

Table 5.1: Mean (standard error) per predictability measure per condition.

	Explicit	Implicit
LS	15.20 (0.53)	16.30 (0.51)
SIV	1.00 (0.03)	1.05 (0.02)
RS	0.00 (0.00)	0.81 (0.15)

Table 5.2: Correlation coefficients between the three predictability measures. * indicates significance at the .05 level.

	LS	SIV
SIV	.07	
RS	-.07	.30*

that the data are not significantly different from the model cannot be rejected, and thus reflects good fit. The paths of a piecewise SEM model can be interpreted as one variable leading to another variable. The coefficients of these paths can be used to calculate indirect effects. To ensure that the coefficient (and standard error) estimates of the separate regression models were stable, we bootstrapped each individual model (10k iterations) using `boot` (Angelo Canty & B. D. Ripley, 2024) and `bootmer` (for mixed-effects models, Bates & Sarkar, 2007). These bootstrapped estimates were then used to obtain indirect and total effects using `semEff` (Murphy, 2022).

5.4.2 Descriptives

A summary of the unexpectedness measures per condition is presented in Table 5.1. Semantic information value (SIV) and relation surprisal (RS), but not the other unexpectedness measures, were found to be correlated (r = .30 [.11, .47], p < .01), as shown in Table 5.2. The piecewise SEM illustrated in Figure 5.1 revealed a good fit (Fisher's C = 4.22, df = 6, p = .65), with no other significant paths as shown by tests of directed separation (p > .33). The bootstrapped estimates of the effect are presented in Table 5.3.

As expected, the presence of a connective significantly predicts the surprisal of the upcoming relation. For SIV and LS, connective presence was not a significant predictor. Although the connective does not directly affect SIV, there is a significant indirect effect of SIV through RS: the connective reduces RS, which in turn reduces SIV. In other words, the connective helps readers to make a more accurate prediction

Table 5.3: Direct and indirect effects of connectives, predictability measures and length on the different predictability measures. * indicates significance at the .05 level. Paths refer to Figure 5.1

	predictor	path	type	β	SE	95% CI	
LS	conn	c	direct	.15	.08	[-.01,.30]	
	length	f	direct	.62	.06	[.48,.72]	*
SIV	conn	b	direct	.02	.11	[-.20,.24]	
	length	e	direct	.27	.08	[-.11,-.41]	*
	RS	d	direct	.32	.08	[.18,.47]	
	conn	ad	indirect	-.16	.04	[-.25,-.09]	*
RS	conn	a	direct	-.50	.05	[-.58,-.41]	*

about the upcoming relation, which in turn helps them to predict the content of the relation more accurately.[4]

According to the Uniform Information Density Hypothesis, relations should only be signaled by a connective if they are unexpected. Readers may therefore expect content that is more surprising when the relation is signaled (i.e. in the explicit condition).[5] If so, the content of `result` continuations in the implicit condition should be more similar to each other than those in the explicit condition. We therefore calculated the average cosine similarity between all `result` continuations for each item and condition. A pairwise t-test revealed no significant difference in average cosine similarity in the explicit (mean = 0.20, SD = 0.06) vs. the implicit (mean = 0.19, SD = 0.06) condition ($t(46) = 1.34$, $p = .19$).

In sum, we indeed find evidence that the connective reduces the unexpectedness of upcoming material, specifically with respect to the upcoming relation and indirectly to the content of that relation. In the next section, we examine whether these factors also influence processing.

[4]It is also possible that readers' predictions about upcoming content are through association, in which case semantic prediction would influence relation prediction rather than vice versa. However, here we assume that the context raises a question-under-discussion that leads to a prediction about the relation type (Clifton Jr & Frazier, 2012; Kehler & Rohde, 2017) and in turn about the content of such a relation.

[5]This would not show up in the analysis above, since the positive relation between relation surprisal and semantic information value might be driven by those continuations in the implicit condition that are not `result` relations.

5.5 Experiment 1: Self-paced reading

The previous section described how we obtained measures of predictability at various levels and that the presence of a connective (indirectly) reduces the unexpectedness of the upcoming relation and content. Crucially, we are interested into what effects these factors have on on-line processing. Does unexpectedness at the relation, semantic and local level disrupt reading? And does the connective facilitate reading directly or indirectly (i.e. by reducing relation surprisal)? To answer these questions, we here present a self-paced reading study.

5.5.1 Methodology

Materials

The 24 items from the pretest were included in the self-paced reading task. A spill-over sentence was added to prevent item-level wrap-up effects. This spill-over sentence was the same across conditions. An example item in the `exp` condition is shown below. Reading times were measured on the target region (here in bold). The items were divided in multi-word chunks, illustrated with forward slashes here, that participants would see one at the time. The target region was never the last chunk of a line.

(40) Angela used to live / in a small flat / in Atlanta. / She didn't pay rent / for months. / She / **was evicted** / by her landlord. / Angela decided / to move to a rural area.

All items and regions were similar in length across conditions. The target region was between 2 and 5 words long (8-22 characters, mean = 15.8) and the spill-over region consisted of 1 to 3 words (8-18 characters, mean = 12.6). The preceding context was on average 18 words long in both predictability conditions (81-133 characters, mean = 105.4), with one less word in the implicit condition. The difference in context length between conditions was maximally 3 words.[6] The experimental items were dispersed with 28 fillers, which contained different coherence relations and connectives. The items were divided across four lists, with each list containing 6 items in every condition. In addition, verification statements were constructed for half of the items.

[6]There is one exception, in which the item was 11 words longer in the item with higher predictability.

Participants

In total, 129 adult native speakers from English, living in the United States, participated in the experiment. They were recruited from Prolific and did not participate in any of the pretests. After removing data from participants who failed the attention check, data from 121 participants (mean age: 37; range: 19-71 years; 54 female) were left for analysis.

Procedure

Participants read the items in a non-cumulative moving window self-paced paradigm, implemented in PCIbex (Schwarz & Zehr, 2021). They first saw three practice trials, before moving on to the 52 trials of the experiment itself. Participants were randomly assigned to one of four lists. The order of the items was pseudo-randomized for every participant, such that no two items in the same condition followed each other. Half of the items were followed by verification statements. Participants were given the opportunity to take a break half-way through the reading task.

Analysis

Data from participants (n=8) who answered less than 70% of the verification statements correctly were removed from analysis.[7] In addition, we removed data from trials (n=6) in which participants spent more than a minute, as this might indicate that they took a break from the experiment. Furthermore, we removed reading times on the target region that were above 2000 ms (n=40) or below 100 ms (n=2), as well as reading times that were more than 2.5 SD away from the participant's mean (n=78), removing 4.1% of the data points for the target region. 2782 data points were left for analysis. The reading times of the region preceding and following the target region were cleaned in a similar manner, excluding 3.6% and 4.00% of reading times for these regions respectively.

Log-transformed reading times of the target region were analyzed in a mixed-effects regression model. As in the analysis above, connective presence was sum-coded (`exp` as -1; `imp` as 1). The model also contained the three predictability measures, as

[7]Responses to four items were removed before this analysis, since either the overall accuracy was very low ($< 60\%$, $n = 3$), indicating that the questions were too difficult, or participants commented that the question was ambiguous.

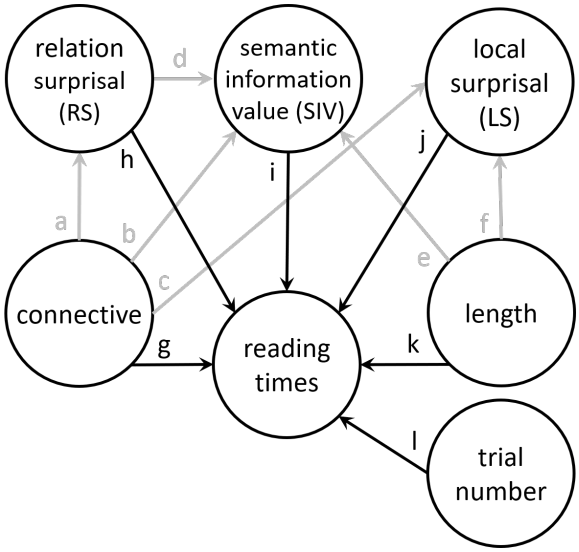

Figure 5.2: Structure of the piecewise structural equation model including reading times with the new paths highlighted.

well as trial number and length in characters.[8] All continuous variables were centered and scaled. We started out with a maximal random effect structure, but removed those random effects that did not improve model fit ($\alpha = .2$), following (Matuschek et al., 2017), to reduce unnecessary complexity of the model.

5.5.2 Results

The raw reading times per region are plotted in Figure 5.3. As can be seen, on average the reading times are higher in the implicit condition. However, these raw means do not take into account variation between participants and items, nor the effects of length, trial and the unexpectedness predictors. The regression model on reading times was added to the piecewise SEM outlined in the previous structure, resulting in the structure illustrated in Figure 5.2. The model again achieved good fit (Fisher's C = 12.24, df = 12, $p = .43$) with no significant tests of directed separation. The bootstrapped estimates of the effects are provided in Table 5.4. Note that adding these paths does not change the estimates of the intermediate paths presented in 5.3.

Our first research question is whether the different measures of predictability influence reading times, and specifically if we can find evidence for relation prediction.

[8]To account for rapid expectation adaptation effects (Fine et al., 2013), we tested whether the effect of the various predictors of interest changes over the course of the experiment in a separate model. The interaction between trial number and the other predictors were not significant.

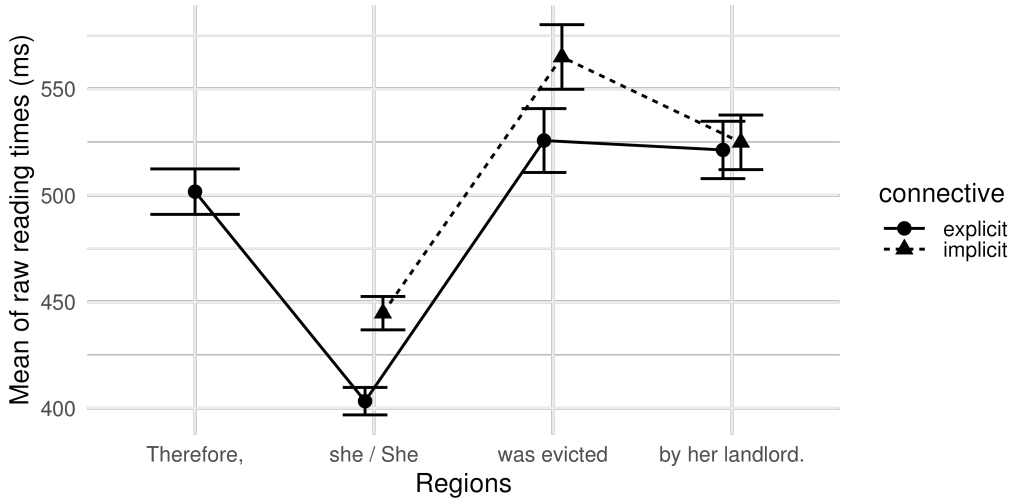

Figure 5.3: Mean reading time per condition per region. Error bars represent 2 standard errors.

The fitted reading times of the linear regression model are visualized in Figure 5.4. With respect to local surprisal, this effect did not reach significance, as shown in Table 5.4. Semantic information value, however, does significantly predict reading times, such that higher SIV is associated with longer reading times. There is no direct effect of relation surprisal on reading times, but in total, higher RS is associated with slower reading due to a significant indirect effect of RS through SIV. As discussed in the previous section, higher relation unexpectedness leads to higher semantic unexpectedness, which in turn influences reading times. In other words, higher relation surprisal leads to longer reading times, because it increases semantic unexpectedness.

Secondly, we focus on the effect of the connective. More specifically, we are interested in whether the effect of the connective facilitates reading through facilitating prediction, and whether there is an additional effect of connective presence. First of all, we replicate earlier studies that the target region is read faster in the presence of a connective. To examine whether this processing benefit can be explained by enhanced prediction, we inspect whether there is an indirect effect of the connective on reading times. Surprisingly, this is not the case, despite the fact that most individual paths (a, d and i) are significant. In other words, we find evidence that the connective reduces the unexpectedness of the upcoming relation and, as a result, of the semantic content as well as that lower unexpectedness of the discourse relation is associated with faster reading by reducing the unexpectedness of the semantic content, but we do not find that the connective facilitates reading by reducing unexpectedness. We thus find no evidence that the facilitative effect of the connective can be attributed

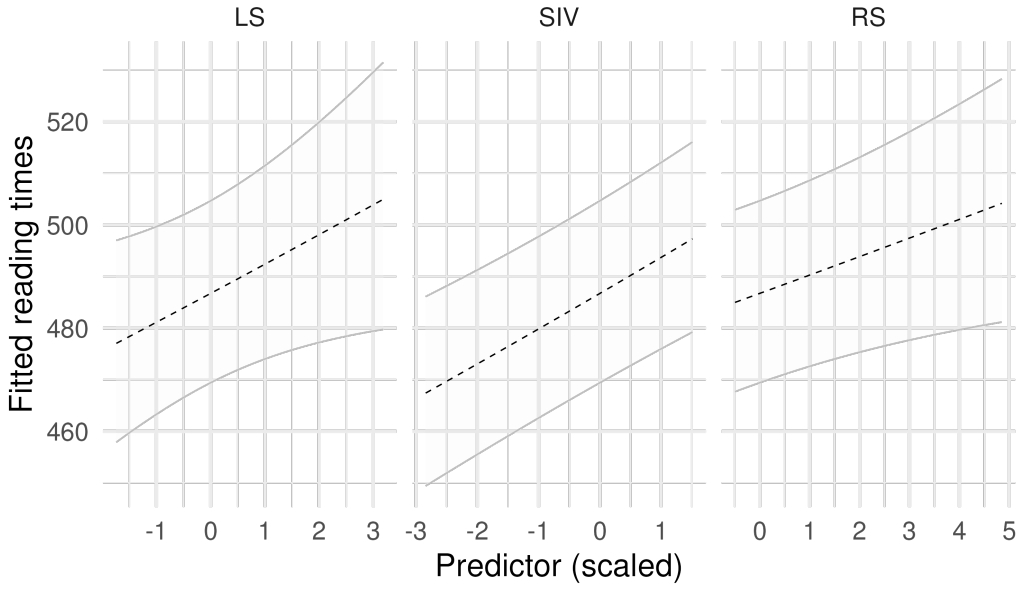

Figure 5.4: The direct effect of the different predictability measures on fitted reading times of the target region.

Table 5.4: Direct, indirect and total effects of the predictors on reading times. * indicates significance at the .05 level. Paths refer to Figure 5.2. Effects of the paths a-f can be found in Table 5.3.

predictor	path	type	β	SE	95% CI	
conn	g	direct	-.06	.01	[-.08,-.03]	*
length	k	direct	.13	.01	[.11,.16]	*
trial	l	direct	-.11	.01	[-.13,-.09]	*
LS	j	direct	.02	.02	[-.01,.05]	
SIV	i	direct	.03	.01	[.01,.05]	*
RS	h	direct	.02	.01	[-.01,.04]	
conn	ah+adi+bi+cj	indirect	-.01	.01	[-.02,.01]	
length	ei+fj	indirect	.02	.01	[.02,.04]	*
RS	di	indirect	.01	.00	[.00,.02]	*
conn	g+ah+adi+bi+cj	total	-.07	.01	[-.09,-.05]	*
length	k+ei+fj	total	.15	.01	[.13,.18]	*
RS	h+di	total	.03	.01	[.01,.05]	*

to more accurate predictions about the upcoming relation type and content. Instead, we find a strong direct effect of the connective, suggesting that there is an additional effect of the connective.

5.5.3 Discussion

The self-paced reading study reveals that different levels of unexpectedness indeed facilitate processing. More specifically, we found evidence that higher relation surprisal and semantic information value predict longer reading times. However, they do not do so independently: relation surprisal leads to shorter reading times through semantic information value. This suggests that relation prediction facilitates reading by allowing readers to make more accurate predictions about the upcoming content.

We replicated earlier findings that the presence of a connective speeds up reading of the clause-initial material following it. Despite finding that semantic and relation prediction facilitate processing, and that the connective reduces relation surprisal (and indirectly semantic information value) as discussed in Section 5.4.1, we did not observe that the connective facilitates reading through the unexpectedness of the discourse relation and content. In other words, we did not find evidence that the connective leads to faster reading times because it enables comprehenders to predict the upcoming relation type and content.

Instead, the presence of a connective directly predicted reading times. There are two possible explanations for this. Firstly, our estimates of the unexpectedness of the relation and semantic content might be inaccurate. This would lead to these measures not capturing differences related to the presence of a connective well, and their effects being attributed to the unexpectedness of the connective directly. We will return to this issue in Section 5.7. The other option is that the facilitating effect of the connective cannot (fully) be explained through reducing unexpectedness. In this scenario, there must be another process that leads to connective facilitation, e.g. propositional integration. Note that self-paced reading measures conflate early and late processes, on which connectives and surprisal might have differential effects. Possibly, the connective also facilitates reading through prediction, but only in early processes. We therefore aimed to replicate these effects in an eye-tracking experiment.

5.6 Experiment 2: Eye-tracking

As outlined above, self-paced reading measures do not provide insights into which stages of processing are influenced by the predictors. More specifically, measures of predictability have been found to affect early processes (cf. De Varda et al., 2023). As discussed in Section 3.5.2, connectives have been found to facilitate both early processes, as shown by shorter first-pass duration and similar measures (Van Silfhout et al., 2014, 2015; Cozijn et al., 2011), but also later processes, reflected in shorter total fixation times and fewer and shorter regressions (Van Silfhout et al., 2015; Cozijn et al., 2011).

We therefore conducted an eye-tracking experiment to distinguish whether these early effects could be attributed to the connective leading to enhanced prediction and the later effects to integration processes. In line with previous literature, we hypothesized that the effect of the connective would show up both in early (first-pass duration) and late measures (total fixation time, regression out probability). More specifically, we predicted the influence of the connective in early processes to (partly) reflect its effect on predictability. This would be reflected in negative effects of relation surprisal and semantic information value (and possibly local surprisal) on reading times (either directly, or indirectly), as well as an indirect of the connective on reading times through these measures of predictability. For late measures, we expect the effect of the connective to mainly be direct, reflecting ease of integration. However, previous research also reveals effects of surprisal on later measures (De Varda et al., 2023; Wilcox et al., 2023; Giulianelli et al., 2023). We therefore also expect that surprisal influences later-pass reading.

5.6.1 Methodology

Participants

79 native speakers of English (age range: 18-36 years; mean age 22 years; 32 female, 1 non-binary) participated in an eye-tracking experiment.[9] They were recruited through adverts at the University of Edinburgh[10] and Saarland University[11] and re-

[9]Due to technical issues, demographic data of 21 participants was lost.

[10]The study was approved by the PPLS ethics committee.

[11]The study was approved by the Deutsche Gesellschaft für Sprachwissenschaft Ethics Committee, see Chapter 2, Section 2.4

ceived monetary compensation. All participants had normal or corrected-to-normal vision.

Materials

The materials consisted of the same 24 items as in the self-paced reading study with minor modifications to adapt to British English. The connective nor the critical region never appeared at the start or the end of a line. The experimental items were combined with 76 fillers that were similar in length, partly from unrelated experiments.

Procedure

Participants were seated at a distance of approximately 60 cm from the monitor and rested their head on a chin-rest. Participants' eye-movements were monitored by an SL Research EyeLink 1000, tracking the participant's right eye. The eye-tracker was calibrated and validated using 11 points. The participant was instructed to read the passages at a natural pace, pressing space bar to continue to the next trial. Two practice trials were displayed. Before presentation of each passage, a fixation mark appeared at the position of the first word of the first sentence. The stories were presented in their entirety on the screen, in a pseudo-randomized, but fixed order. After reading half of the items, participants took a short break from reading and performed two other tasks,[12] before continuing with the eye-tracking experiment. To encourage reading for meaning, participants were presented with a verification statement about story content following 25% of the items. On average, the whole session took approximately an hour.

Analysis

The data was manually reviewed for vertical drift correction and automatically cleaned in DataViewer, such that fixations shorter than 80ms were merged with fixations within 0.5 degrees. Subsequently, fixations shorter than 80ms were removed. Data from one participant was excluded from further data analysis due to low ($< 70\%$) accuracy on the verification statements. In addition, data from 32 individual trials was removed due to low data quality.

[12]These tasks consisted of a short version of the Raven's Progressive Matrices Test and the Author Recognition Task.

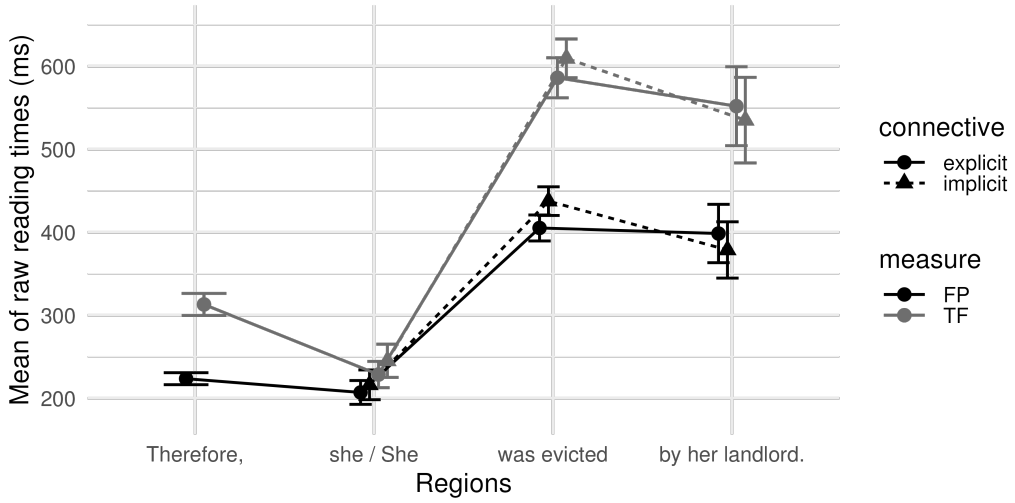

Figure 5.5: Mean reading time per condition per region. Error bars represent 2 standard errors.

Two reading measures were computed: first pass duration (FP) and total fixation duration (TF). First pass duration is the time spent on a non-skipped region before exiting the region in any direction. Total fixation duration is the sum of all fixations on a region, regardless of whether this is during first pass or during a regression to that region. We excluded reading times that were more than 2.5 SD away from the participants' mean per region (for the critical region, this consisted of 0.89% of FP and 2.48% of TF). The procedure for analyzing the reading measures was similar to that of the self-paced reading times, described in Section 5.5.1.

5.6.2 Results

The first-pass duration and total fixation duration per condition and region are visualized in Figure 5.5. When adding the reading models to the piecewise SEM introduced in Section 5.4.2, there was a significant test of directed separation. Unintentionally, trial number significantly predicted local surprisal.[13] We therefore included trial number into the structural equation model, yielding the model structure illustrated in Figure 5.6. Note that this modification did not qualitatively change the relation between the connective and the different predictability measures as described in Section 5.4.2. This updated model yielded good fit for both measures (Fisher's C = 8.85, df = 10, $p = .55$) with no significant tests of directed separation. The estimated effects can be found in Table 5.5.

[13]Note that the item order for the eye-tracking, but not for the self-paced reading study was fixed.

Table 5.5: Direct, indirect and total effects of the predictors on **first-pass duration** and **total fixation duration**. * indicates significance at the .05 level. Paths refer to Figure 5.6.

			FP			TF		
predictor	path	type	β	95% CI		β	95% CI	
conn	g	direct	-.10	[-.15,-.05]	*	-.06	[-.10,-.01]	*
length	k	direct	.18	[.13,.25]	*	.23	[.16,.27]	*
trial	l	direct	-.06	[-.09,.01]		-.02	[-.10,.01]	
LS	j	direct	.10	[.02,.15]	*	.07	[.03,.15]	*
SIV	i	direct	.08	[.04,.13]	*	.14	[.10,.19]	*
RS	h	direct	-.07	[-.13,-.03]	*	-.02	[-.07,.03]	
conn	ah,adi,bi,cj	indirect	.04	[.01,.08]	*	.00	[-.04,.04]	
length	ei,fj	indirect	.09	[.03,.13]	*	.09	[.05,.14]	*
trial	mj	indirect	.03	[.00,.05]	*	.02	[.00,.04]	*
RS	di	indirect	.03	[.01,.05]	*	.05	[.02,.08]	*
conn	g,ah,adi,bi,cj	total	-.06	[-.10,-.01]	*	-.06	[-.10,-.01]	*
length	k,ei,fj	total	.26	[.22,.32]	*	.32	[.26,.36]	*
trial	l,mj	total	-.04	[-.07,.03]		-.00	[-.07,.03]	
RS	h,di	total	-.05	[-.10,-.00]	*	.03	[-.04,.08]	

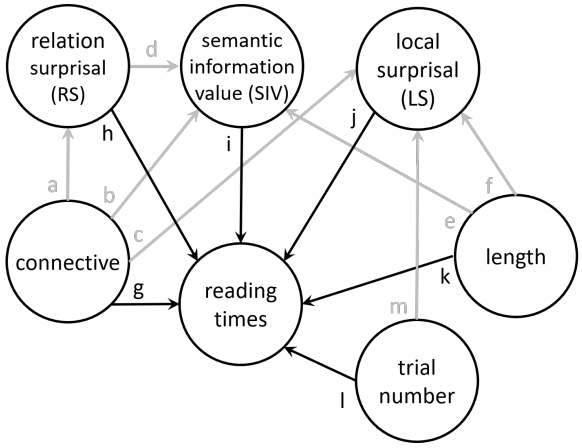

Figure 5.6: Structure of the piecewise structural equation model for the eye-tracking data. Paths for the reading measures are highlighted.

First, we examine whether we find evidence that the different measures of predictability affect the reading measures. For both reading measures, we find a significant effect of local surprisal: as illustrated in Figure 5.7, higher LS leads to longer reading times. Note that this could theoretically also be partly attributed to the order of presentation: there is a significant indirect positive effect of trial number on reading times. However, this is contrary to what is mostly found for the effect of trial on reading and is also the trend for the direct effect of trial: that participants read faster over the course of an experiment. This suggests that LS affects reading times, despite an effect of trial number.

Furthermore, for both reading measures, the finding from the self-paced reading study that high semantic information value predicts longer reading times is replicated (see also Figure 5.7). Also in line with Experiment 1, there is again a positive indirect effect of relation surprisal, suggesting that the positive effect of relation prediction on content prediction influences both first-pass and total fixation duration. Contrary to Experiment 1, however, we also find a significant direct effect of relation surprisal for first-pass duration. As can be seen in Figure 5.7, this is in the opposite direction from what was predicted: When taking into account the effect of other measures of predictability, *lower* relation surprisal predicts longer first-pass duration. In other words, when relation prediction does not facilitate semantic prediction, it slows readers down.

With respect to the facilitating effect of the connective, we find that the connective influences first-pass duration as well as total fixation duration. For first-pass duration, we find an indirect effect, but in the opposite direction from what we hypothesized: The effect of a connective across the different measures of unexpectedness leads to

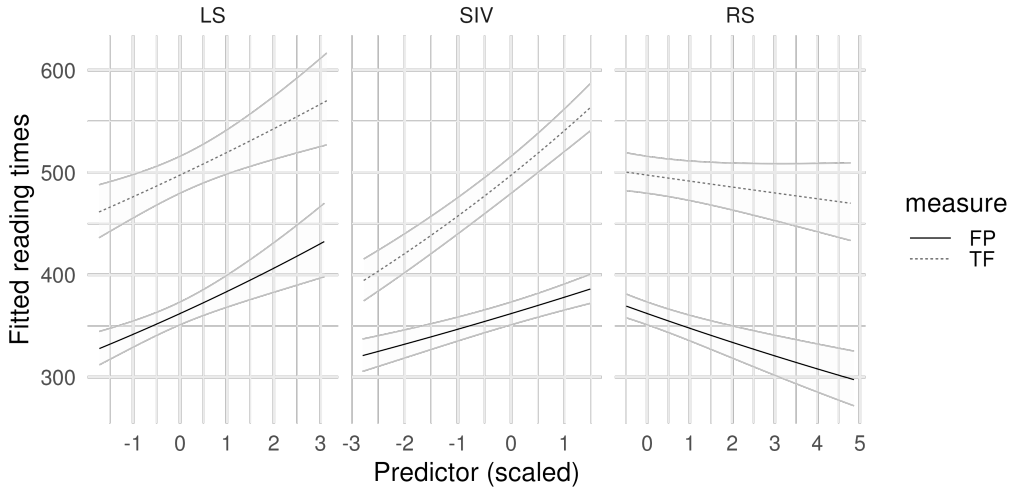

Figure 5.7: The direct effects of the different predictability measures on fitted **first-pass duration** (FP) and **total fixation duration** (TF) of the target region.

slower reading. This is due to the negative direct effect of relation surprisal on reading times, since all other paths are in the expected direction. As in the self-paced reading experiment, we also find a direct effect of the connective for both reading measures. This suggests that the effect of the connective cannot be explained by enhanced prediction. On the contrary, we find that the connective inhibits reading through predictability, at least in first-pass duration, to the negative effect of relation surprisal on reading times.

5.6.3 Discussion

The goal of the eye-tracking experiment was to replicate the findings in the self-paced reading study and examine if there were differences in the effects on early and late processes, as indicated by first-pass duration and total fixation duration respectively. In general, the effects were fairly consistent across the two measures. We found significant direct effects of local surprisal on both eye-tracking reading measures, such that material that was more unexpected was read slower. In addition, for both these measures, content that is more predictable was read faster, as shown by a significant direct effect of semantic information value. We also found an indirect effect of relation surprisal on reading times through semantic information value for both measures, suggesting that relation types that are more predictable help readers to make more accurate predictions about the content, which in turn facilitates reading. Surprisingly,

however, this does not mean that relations that are more expected are read faster. When controlling for effects of semantic unexpectedness, more unexpected relation types led to *shorter* first-pass duration.

How can this positive effect of relation unexpectedness on reading times be explained? One possibility is that readers might expect a more informative (and thus more unexpected) continuation when the relation type is highly predictable, especially when it is marked by a connective (cf. Rohde et al., 2022). This would suggest that the unexpectedness of the content slows down reading less when it is marked with a connective. However, a post-hoc analysis showed no interaction between the presence of a connective and content unexpectedness (β = 0.01, SE = 0.01, t = 0.83, p = .41).[14] In addition, in the pretest (Section 5.4), we found no evidence that readers' off-line expectations about the content of signaled `result` relations is less predictable. Another option is that predictions about relation types only facilitate reading when the content of the relation is also predicted correctly (cf. the indirect effect of relation unexpectedness on reading times). If the content is not in line with what was predicted, readers might need to recompute if this unexpected content is still in line with the predicted relation type. This would then result in longer reading times, even when the relation type is predictable. However, further research is needed to examine this hypothesis.

With respect to the connective, we replicated the finding from the self-paced reading study and earlier work showing that the connective facilitates processing of the material directly following it. However, contrary to our expectations, we did not find that this effect could be attributed to facilitated prediction. In fact, for first-pass reading, the effect of the connective through the predictability measures was even positive, suggesting that the effect that the presence of a connective has on the unexpectedness of the relation slows reading down. As in the self-paced reading experiment, the facilitating effect of the connective was independent from predictability effects. We will discuss the interpretation of this finding in the next section.

5.7 General discussion

This study set out to examine the role of prediction in the processing of discourse relations. Specifically, we investigated whether the unexpectedness of the relation

[14]Model formula: `log(rt) ~ conn * SIV + LS + RS + trial + length + (1 + LS + trial | subj) + (0 + conn + length | item)`

type as well as its content influences the reading of `result` relations. The unexpectedness of the relation type in each item and condition was operationalized as relation surprisal, the negative log probability of the relation type given the context. This was obtained from continuations provided by participants in a pretest. From this pretest, we also estimated the unexpectedness of the content of the relation, by calculating the semantic information value (the average semantic distance between the provided continuations and the target, as in Giulianelli et al., 2023). In addition, we controlled for lexical-syntactic unexpectedness, estimated with empty-context GPT2 surprisal.

In a subsequent self-paced reading and an eye-tracking experiment, we examined whether these measures predict the reading times of `result` relations. We hypothesized that lower unexpectedness would lead to shorter reading times. Across all reading measures, we indeed found that content that was less unexpected was read faster. This content was easier to predict when the relation type was also less unexpected. As a result, reading times on all measures were faster when the relation type could be predicted, but only when this resulted in more accurate predictions about the content. In other words, there was a positive indirect effect of relation surprisal through semantic information value. However, when controlling for the effect of semantic unexpectedness, `result` relations were read slower when this relation type is more expected. Possibly, readers confirm their prediction about the relation type when the content is different from what they expected. However, further research is needed to replicate this effect and investigate this hypothesis.

These findings also suggest that differences in processing difficulty between relation types (e.g. Sanders & Noordman, 2000; Xu et al., 2015, see also Chapter 2, Section 2.5.2) cannot be explained by the unexpectedness of these relation types, but may rather be due to differences in cognitive complexity or default interpretations (Sanders, 2005). Although some relation types are more expected than others and these relations also require more processing difficulty (cf. Mulder, 2008), we do not find evidence that the unexpectedness of the relation type independently inhibits reading. Instead, we find that `result` relations were read *slower* when this relation type was more expected, at least in first-pass duration. However, we do find that the content of more unexpected relation types is more difficult to predict and hence relation unexpectedness inhibits reading in some cases. Possibly, differences between relation types can partly be explained by differences in the unexpectedness of the content of such relations. The content of `concession` relations might be more difficult to predict than of `result` relations.

A second goal of this study was to investigate whether the facilitating effect of the connective on processing is due to the connective enabling readers to make more accurate predictions about the following relation type and semantic content. The pretest shows that the presence of a connective indeed reduces the unexpectedness of the relation type and in doing so, also that of the semantic content. In addition, we replicate earlier findings that the presence of a connective facilitates clause-initial reading (e.g. Cozijn et al., 2011; Van Silfhout et al., 2014). However, we did not find evidence that the facilitative effect of the connective could be attributed to the connective reducing the unexpectedness of the upcoming material. On the contrary, for first-pass duration we found that the effect of the connective on the unexpectedness of the relation type slowed readers down, due to more expected relations being read slower, as discussed above. Instead, the facilitating effect of the connective was independent from effects of unexpectedness.

Our findings thus suggest that connectives provide more 'processing instructions' (cf. Van Silfhout et al., 2015) than updating predictions on the upcoming relation type and content. What then constitutes the facilitating effect of the connective? Possibly, propositional integration is a purely bottom-up process that is not influenced by expectations about the relation type from the preceding context. Establishing the discourse relation is simply easier in the presence of a connective as linguistic input on the type of relation. Another possibility is that the presence of a connective influences readers' processing strategy. Indeed, connectives have been shown to trigger more, but shorter, regressions (Van Silfhout et al., 2015). With respect to the reading times of the relation itself, readers might process the material following the connective more *shallowly*, only to process the relation more *deeply* at sentence wrap-up, reflected in longer reading at this region (Millis & Just, 1994; Cozijn et al., 2011). However, further research is needed to explore these suggestions. Here we show that the facilitating effect of the connective cannot fully be explained by differences in unexpectedness.

Finally, we examined whether readers might expect less predictable content in relations signaled with a connective. However, we did not find any evidence for this. In the pretest, the content of `result` continuations was similar in the explicit and implicit condition. In addition, the effect of the unexpectedness of the content on reading times was not significantly modified by the presence of a connective.

Our findings were consistent across reading times from self-paced reading and early and late eye-tracking measures, which strengthens our findings. The only differences were that the surprising negative effect of relation unexpectedness was only found for

first-pass duration, which also resulted in a positive indirect effect of the connective for this measure. This effect should be replicated in future work. In addition, the effect of local surprisal was only found to be significant in the eye-tracking experiment. Note that this measure of unexpectedness did not take into account the preceding content, contrary to the other measures of unexpectedness. Possibly, contextual effects are more pronounced in self-paced reading due to the incremental presentation of the stimuli (Koornneef, 2021).

5.7.1 Limitations

There are several limitations to our study. First of all, we rely on the assumption that our estimates of the unexpectedness of the relation type and the semantic content are accurate. If they are not, their effect on processing difficulty might be attributed to other factors in the model (e.g. the connective). There are several factors that might influence the accuracy of our unexpectedness estimates. For example, readers' predictions generated in off-line tasks might be different from those generated in on-line tasks (cf. Staub et al., 2015). In addition, to estimate the unexpectedness of the type and content of the relation, we sampled from a limited number (n=20) of human continuations. The accuracy of these measures likely increases with sample size (cf. Giulianelli et al., 2023). Furthermore, we used a large language model to estimate the semantic similarity of those continuations to our target continuations, but it is unclear how similar this is to human similarity measures. Finally, if the language model or participants generating unexpectedness measures are not adapted to the readers in the processing study, these estimates will be less accurate (cf. Škrjanec et al., 2023). Nevertheless, the fact that these measures captured some variation in reading times and the consistency across different reading measures suggests that they are somewhat accurate. Further research, possibly with different measures of predictability or larger sample sizes both for obtaining the estimates and for assessing processing difficulty, should validate the findings of this study.

A second limitation is that we only investigated a single relation type: `result` relations. Effects of the unexpectedness of the relation type might be more pronounced when examining a wider variety of relation types. Furthermore, the predictability of `result` relations might be especially high, because readers have been argued to infer causal relations by default (Sanders, 2005). To combat this, we ensured variation in the predictability of our items (see Appendix D). In addition, we chose to investigate a single relation type to keep factors such as cognitive complexity constant. Further research could examine whether the effect of the unexpectedness of the relation type

is stronger when considering other relation types and whether this factor can explain differences in processing difficulty between relation types.

Finally, we acknowledge that reading times do not allow for measuring prediction processes directly, as processing difficulty is measured when the material is presented. Effects of unexpectedness could thus also reflect other processes, such as integration difficulty (Wong et al., 2024). We consider our study as a first investigation on the role of prediction in discourse relation processing during natural reading.

5.8 Conclusion

In this chapter, we examined whether readers' sensitivity to linguistic signals depends on characteristics of the discourse relation (Research Goal 2 in Chapter 1). More specifically, we investigated how the facilitative effect of the connective relates to the predictability of the discourse relation. We hypothesized that the connective enables readers to make more accurate predictions about the upcoming discourse relation type and the content of the second argument, which would facilitate reading that argument. The pretest showed that the relation type is indeed more predictable in the presence of a connective, which in turn allows for more accurate predictions about the content. The relation was read faster when preceded by a connective, replicating findings from earlier studies (see also Chapter 4, Experiment 2), but the facilitating effect of the connective on on-line processing was independent from effects of predictability in all reading measures. We thus find no evidence that readers are more sensitive to the presence of a connective when the relation type is less predictable. In the next chapter, we examine whether the processing of `result` relations is also facilitated when signaled by a non-connective cue.

Data availability

All items, data and analysis scripts can be found here: `https://osf.io/ey52b/`

Funding

This work was supported by the German Research Foundation (DFG) under Grant 232722074 - SFB 1102 ("Information Density and Linguistic Encoding").

Dissemination

This research is based on the following publications:

- **Marchal, M.,** Scholman, M.C.J., Sanders, T.J.M., & Demberg, V. (2024). What processing instructions do connectives provide? Modeling the facilitative effect of the connective. In *Proceedings of the 46th Annual Meeting of the Cognitive Science Society* (pp. 3435–3441).
- **Marchal, M.**, Scholman, M.C.J., Sanders, T.J.M. & Demberg, V. (in prep.). Predicting discourse relations: Understanding the facilitating effect of the connective.

In addition, (parts of) this research have been presented at the following conferences:

- **Marchal, M.**, Scholman, M.C.J., Sanders, T.J.M., & Demberg, V. (2024). Predicting discourse relations: The processing benefit of a connective. *Poster at Architectures and Mechanisms for Language Processing, Edinburgh, Scotland, 5-7 September.*
- **Marchal, M.**, Scholman, M.C.J., Sanders, T.J.M., & Demberg, V. (2024). What processing instructions do connectives provide? Modeling the facilitative effect of the connective. *Poster at the Annual Meeting of the Cognitive Science Society, Rotterdam, the Netherlands, 24-27 July.*
- **Marchal, M.**, Scholman, M.C.J., Sanders, T.J.M., & Demberg, V. (2024). What processing instructions do connectives provide? Modeling the facilitative effect of the connective. *Talk at Society for Text and Discourse, Chicago, Illinois, 17-19 July.*

- **Marchal, M.**, Scholman, M.C.J., Sanders, T.J.M., & Demberg, V. (2023). What processing instructions do connectives provide? Disentangling relation and content prediction. *Poster at LingCologne 'Prediction in Language'*, Cologne, Germany, 16-17 June.

Part III

Beyond connectives

Chapter 6

Beyond words
Clause structure as a cue for discourse relations

The previous chapters examined the role of the connective in reading, showing that they facilitate on-line processing. In this chapter, we expand this line of research to other discourse relation signals, since much less is known about how these signals influence the processing and representation of discourse relations (cf. Chapter 3). More specifically, we investigate readers' sensitivity to a non-lexical cue: clause structure. The primary function of this cue is not in agreement with that of the discourse relation it signals (cf. Chapter 3). They can therefore only be acquired through statistical information. These systematic correlations have been argued to influence phonological, syntactic and lexical processing, but it is still unknown to what extent these correlations influence discourse-level processing. We address this question by examining whether clause type serves as a cue for discourse relations. A continuation task shows that readers' expectations about upcoming discourse relations is influenced by the co-occurrence of gerund free adjuncts and specific discourse relations found in natural language. However, we did not find evidence that clause structure facilitates the on-line processing of these discourse relations in a self-paced reading task, nor that readers have a preference for these relations in a paraphrase selection task. The present research extends previous research on discourse relation processing, which mostly focused on lexical cues, by examining the role of non-semantic cues. We show that readers are aware of correlations between clause structure and discourse relations

in natural language, but that, unlike what has been found for lexical cues, this information does not seem to influence on-line processing and discourse interpretation. Thus, readers' sensitivity to linguistic signals in establishing coherence depends on characteristics of the signal itself.[1]

[1]This chapter is based on (and in parts identical to) the following publication: **Marchal, M.**, Scholman, M.C.J., & Demberg, V. (2023). How statistical correlations influence discourse-level processing: Clause type as a cue for discourse relations. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *50*(5), 796-807. doi:10.1037/xlm0001270

6.1 Introduction

Lexical-semantic and syntactic-structural information are important sources of information for constructing meaning at the sentence level. Syntactic structure reveals who did what to whom and when, while lexical-semantic elements convey more information about the identity of these entities. At the discourse level, lexical-semantic information has also been shown to play a crucial role. As discussed in Chapter 3, connectives provide 'processing instructions' on how the reader should relate the information in different propositions. Much less is known about which other types of information from the text inform the reader how to construct discourse relations between textual elements. In the present study, we focus on whether and how non-lexical information helps readers establish relations between parts of the text. More specifically, we examine how statistical correlations between a linguistic phenomenon (i.e. clause structure) and discourse relations influence various aspects of discourse processing and comprehension.

The information that readers obtain from a linguistic signal is not always deterministic, but comprehenders still need to construct meaning from it. Words, but also utterances, are often ambiguous or polysemous, yielding multiple possible interpretations. In some cases, the intended meaning can be deduced from the context. If such contextual information is not available or insufficient, readers are able to exploit the probabilistic meaning of lexical items. For example, Asr & Demberg (2020) show that readers' interpretation of the ambiguous connectives *but* and *although* corresponds to their distributional meaning in production, where *although* is more strongly associated with a concessive meaning compared to *but*. Thus, comprehenders rely on statistical correlations between lexical elements and the discourse structure to infer meaning from text.

To obtain a better understanding of how readers construct meaning from the linguistic signal, we need to gain more insight into which linguistic information they take into account when processing a text. Beyond lexical items, there are many other linguistic signals, including syntactic information, that can correlate with a specific meaning or interpretation in natural language. These signal∼meaning correlations can enrich comprehenders' language representation and help them to construct meaning from text. We hypothesize that readers' sensitivity to these correlations influences their processing similarly to lexical information: the cues can affect how a text is interpreted and elicit expectations about upcoming material.

To illustrate, tense has been shown to co-occur frequently with specific chronological ordering of events. Two subsequent clauses in simple past tense often describe subsequent events, as in (41a) below, though the events can also happen simultaneously (see (41b)).

(41) a. Jade picked up some groceries. She visited her grandmother.
b. Jade picked up some groceries. She saw an old friend.

Readers can use this statistical information about the relation between tense and the temporal order of the events to create expectations about upcoming events or to interpret how the events are structured in time. Whether comprehenders are indeed sensitive to this type of information is an open question, however. Statistical regularities have been shown to influence phonological, lexical and syntactic processing (cf. Saffran et al., 1996; Savic et al., 2022; Maye et al., 2002), but it is not yet clear whether readers are also sensitive to co-occurrence at the discourse level.

The present study sets out to investigate to what extent non-lexical cues can affect discourse-level processing, by examining how clause structure influences readers' processing of discourse relations. The clause structure we focus on are gerund free adjuncts (GFAs). These are subordinate clauses that start with a present participle, as in (42) (GFA between square brackets).

(42) [Walking to school,] the girl sang a song.

We hypothesize that it is particularly the information provided by the structure of the clause (encoded in the present participle) that yields information about upcoming discourse relations due to its co-occurrence with specific discourse relations, independent of semantic cues.

We examine the effect of this clause structure on several aspects of discourse relation processing. In the next section, we will first review previous research on discourse relational cues and their influence on discourse processing, which inform our research questions and hypotheses in the following section. Section 6.4 presents a corpus investigation, which examined the co-occurrences of discourse relations with this specific clause type. The corpus data revealed a specific distribution of discourse relations that GFAs occur with in natural language, distinct from the relational distribution of non-GFAs. Our findings show that GFA often co-occur with causal relations, corroborating previous studies suggesting that this clause type signals `result` relations (Danlos et al., 2018; Hoek & Scholman, 2023). Sections 6.5 through 6.7 present a series of experiments following up on these findings, which investigate several aspects of discourse processing and comprehension.

6.2 Background

6.2.1 The effect of discourse relational cues

Various types of discourse relational cues have been shown to influence discourse processing and representation. One of the most well-studied type of cues is connectives (see also Chapter 3, Section 3.5.1 and 3.5.2). However, as discussed in Chapter 3 (Section 3.6), connectives are highly informative, specialized, immutable lexical signals for discourse relations. Non-connective signals do not adhere to these features and might therefore be less likely to facilitate processing, since they are less salient and informative. Indeed, findings on the role of non-connective signals for discourse relations are mixed. Crible (2021) found that the presence of negation in the first argument cancels out processing difficulty associated with `concession` relations. However, Grisot & Blochowiak (2017) find no evidence that tense, which was shown to be a signal of temporal relations in corpus research, facilitates reading of such relations.

Note that the cues investigated by Grisot & Blochowiak (2017) and Crible (2021) were conceptually related to the discourse relation: tense provides information about the temporal position of the individual clause and `concession` relations are often referred to as *negative* causal relations. In other words, these cues show agreement (cf. Chapter 3, Section 3.3.4). This is not the case for collocational cues, which solely provide information about discourse relations through their co-occurrence. It is unclear how such cues influence the processing of discourse relation. GFAs, on the other hand, do not provide any tense, mood or aspect information (Kortmann, 2013).

Crible & Pickering (2020) also investigate a non-lexical cue that is not in agreement with the discourse relation. They find that lexical-syntactic parallelism can serve as a cue for `contrast` relations, facilitating the processing of such relations. Note, however, that this effect was only found when participants were explicitly primed to disambiguate the discourse relation. This suggests that non-lexical information might only influence processing when readers are forced to process the text deeply.

All three studies discussed above (Grisot & Blochowiak, 2017; Crible, 2021; Crible & Pickering, 2020) tested the effect of discourse relational cues in interaction with connectives. However, non-connective cues might be especially salient when no specialized signal for the discourse relation is available. Contrary to these previous studies, the present study thus focuses on whether readers use purely statistical information when processing discourse relations in the absence of connectives.

6.2.2 Statistical correlations in language processing

In the present chapter, we examine the influence of non-connective cues from the perspective of a statistical learning account. Tracking regularities in language has been argued to enable a wide range of language processing: from word segmentation (Saffran et al., 1996), to the formation of phonetic (Maye et al., 2002) as well as syntactic categories (Gerken et al., 2005). These abilities are often tested in artificial settings, but psycholinguistic research has shown that language users' ability to extract statistical information from implicit learning tasks is strongly correlated with the extent to which they are able to predict words from contexts (Conway et al., 2010; Misyak & Christiansen, 2012). With respect to semantics, linguistic distributional models have been shown to accurately predict human semantic priming (Günther et al., 2016; Lund et al., 1995; Mandera et al., 2017). Moreover, word co-occurrence plays an important role in word learning, both for children and adults (Unger et al., 2020; Savic et al., 2022). However, previous studies have mostly focused on lexical semantics rather than meaning at the discourse level.

There is limited evidence that statistical correlations also influence language processing at the discourse level. For example, Arnold and colleagues found that the interpretation of pronouns is affected by frequencies of certain referential patterns. In natural language, pronouns often refer to the subject of the previous sentence. Participants are more likely to interpret ambiguous pronouns as referring to the subject of the previous sentence, when they have had more exposure to this pattern, either as more language experience in general (Arnold et al., 2018), as well as by recent exposure (Johnson & Arnold, 2023). This shows that language users are also sensitive to statistical correlations at the discourse level.

With respect to discourse relations, readers have also been shown to exploit the co-occurrence of discourse connectives and relation senses to infer their meaning. For example, the connective *but* can indicate a contrastive (as in 43a) as well as a concessive (as in 43b) relation between clauses.

(43) a. Alex wanted ice cream for dessert, but Ike wanted tiramisu.
b. Alex wanted ice cream for dessert, but he ordered tiramisu.

In natural language, *but* is more strongly associated with a contrastive meaning compared to *although*. Asr & Demberg (2020) show that this is reflected in readers' interpretation of these two ambiguous connectives. This suggests a probabilistic account of the interpretation of relational cues, in which readers track statistical correlations between relational cues and their meaning in order to infer its meaning.

Language users do not only keep track of the occurrence of discourse relations with lexical items, but also with more general linguistic structures. `Reason` relations, as in (44b), are much less frequent than `result` relations, as in (44b), in Korean monologues (e.g. written texts) compared to English (Yi & Koenig, 2021), as backward causal relations are rarely encoded with a connective in Korean.

(44) a. Peter praised Linda. She had won the race.
b. Linda had won the race. Peter praised her.

Yi & Koenig (2021) show that speakers of Korean, who encounter `explanation` relations rarely, produce this relation type less often than speakers of English, who have more experience with both types of relations, in an experimental setting. This finding could not be attributed to cultural differences, as the effect disappeared in a genre (dialogues) in which there was no difference in the frequency of the two types of relations.

To sum up, language users track statistical patterns at various levels of language and use this information to infer meaning (e.g. about pronouns or discourse connectives). The present study examines whether language users are also sensitive to this information at a more abstract level: the co-occurrence of grammatical structure and discourse relations.

6.3 The present study

The research question that this chapter addresses is whether clause structure, specifically GFAs, also serves as a cue for discourse relations. Contrary to previous studies, we examine a non-lexical cue that does not carry any semantic information. The only information about the discourse relation that readers can obtain from this cue is its statistical correlation with certain discourse relations in natural language. This study will therefore provide more insight into the type of information that readers take into account when processing language at the discourse level.

We will examine this question from various perspectives in order to gain a comprehensive understanding of which aspects of discourse processing are influenced by statistical correlates. First, we present corpus research, which shows the specific discourse relations that this clause type correlates with. The findings of the corpus study will generate the hypotheses for a series of experimental studies investigating the role of clause structure on various aspects of discourse processing. In Section 6.5, we examine whether the discourse relations that occur with this type of clause

structure guide readers' expectations in an off-line production task. In Section 6.6, we examine whether the patterns found in natural language are also reflected in on-line processing, using a self-paced reading task. Finally, in Section 6.7, we investigate how statistical information affects discourse representation. This will be examined in a paraphrase selection task, contrasting various relation senses.

In sum, the present research will allow us to investigate (a) whether the distribution of discourse relations in natural language is dependent on clause structure, as well as whether clause structure (b) guides readers' off-line expectations for discourse relations, (c) facilitates the on-line processing of expected relations and (d) influences readers' interpretation of discourse relations.

6.4 Corpus study

To examine the role of GFAs as a discourse relational cue, we first explored the occurrences of GFAs in discourse relations in corpora. Should clause structure serve as a cue for a specific distribution of discourse relations, the distribution of discourse relations should be different for clauses containing a gerund free adjunct than for full matrix clauses (cf. the *information criterion* in Chapter 3, Section 3.2). In addition, this should be the case regardless of any semantic cues in the segments (i.e. connectives or alternative lexicalization). We therefore investigated the distribution of discourse relations containing a GFA and compared it to the relations that occur in inter-sentential implicit discourse relations.

6.4.1 Methodology

Prior work has shown that the interpretation and distribution of free adjuncts is heavily dependent on genre (Kortmann, 2013). We therefore include data from two corpora in our dataset: the PDTB 3 (Webber et al., 2019) and the Blog Authorship Corpus (BAC, Schler et al., 2006). The PDTB is a discourse-annotated corpus of newspaper text that contains annotations of full clauses as well as free adjuncts. We selected those free adjuncts from the PDTB3 that were headed by a participle ending in *-ing*, and included only those which had been annotated as either an implicit or alternative lexicalization relation (see below).[2] In addition, we sourced GFAs from the

[2]Note that in the PDTB, the free adjunct is always coded as the second argument, regardless of its position in the sentence (sentence-initial or -final). This affects the directional order of the relation sense (e.g., REASON versus RESULT). We therefore recoded the relation sense of sentence-

Blog Authorship Corpus, which consists of blog posts from 2004. The corpus contains a wide variety of text types, including diary-style text, narratives and instructions. The first 4,750 files of the corpus were parsed using SpaCy to find clauses starting with a participle ending in *ing*. We further manually identified GFAs in the corpus. A more elaborate description of the corpus methodology can be found in the Online Appendix.

In total, 619 GFAs were identified in this subset from the blog corpus. The relation sense between the GFA and its matrix clause was annotated using the PDTB 3 annotation guidelines, with the exception of identifying alternative lexicalizations and the `+Belief` and `+SpeechAct` features. Inter-annotator agreement on 10% of the relation senses was moderate (k=.63), but comparable with other implicit annotation efforts (Hoek et al., 2021c). In some cases, the meaning of the present participle indicated the discourse relation between the GFA and the matrix clause. For example, in (45), the verb *using* signals the discourse relation `arg1-as-manner`.

(45) [*Using* small electrical shocks applied to her feet,] they were able to monitor sensory nerves.

We categorized these cases as alternative lexicalizations and excluded them from further analyses, since we were interested in the role of GFAs as a structural cue for discourse relations.[3] We identified 347 cases (26.9%) of alternative lexicalizations in the PDTB and 143 instances (23.1%) in the BAC. These cases were removed from subsequent analyses.

6.4.2 Findings

In total, we analysed 1,418 GFAs. Of these, 944 come from the PDTB 3 and 474 from the BAC. GFAs occurred with a high number of relation senses: 20 out of 29 possible relation senses were identified in the corpora. The relation senses varied widely in their frequency. As can be seen in Table 6.1, `result` relations were the most frequent, occurring in almost 25% of the cases. This type of relation, in which the consequence follows the cause, is illustrated in 46 and 47 below (GFA between squared brackets).

initial GFAs to make it more comparable with the annotation of implicit full clauses, where the first clause in the text is also the first argument.

[3]Note that we take a slightly different approach to alternative lexicalizations from the PDTB to ensure consistency across annotations and the different corpora. We therefore re-annotated the items for their alternative lexicalization in the PDTB corpus to ensure consistency across annotations and the different corpora. More details can be found in the online appendix.

(46) [Feeling they should devote more time to their families or their jobs,] many yuppies are skipping their once-sacred workout.

(47) Second, they channel monthly mortgage payments into semiannual payments, [reducing the administrative burden on investors].

Synchronous relations, where the two events would occur simultaneously, without a causal relation, also occurred often in the corpus, followed by `arg2-as-detail`, where the second clause provides more detailed information about the first clause. These relations are illustrated in (48) and (49).

(48) [Praying to the over-the-counter drug gods that the medicine too would go straight to my brain,] I began looking for my keys.

(49) The House voted to boost the federal minimum wage for the first time since early 1981, [casting a solid 382-37 vote for a compromise measure backed by President Bush].

The majority of the GFAs (83.6%) occurred in non-initial position, as in (49) above. This was the case in both corpora. In the current study, we are most interested in sentence-initial GFAs, given that we aim to study whether GFAs can elicit discourse expectations, and such expectations are optimally elicited when the GFA occurs early in the sentence. We therefore looked at whether the position of the GFA influences the relational distribution. Overall, the distributions are fairly similar. The three most-frequent relations with sentence-initial GFAs are also the most frequent relations with non-initial GFAs. Note that for `result` this means that the GFA contains the cause if it is in sentence-initial position, as in (46), but the consequence if it is sentence-final, as in (47). However, within the subset of initial GFAs (n=233), `synchronous` relations are slightly more frequent than `result` relations (cf. Table 6.1). This suggests that clause structure could also be argued to be a cue for `synchronous` relations. However, this interpretation is also influenced by other factors, such as verb tense (cf. Grisot & Blochowiak, 2021). More specifically, the combination of the present participle with a specific tense in the matrix clause can lead to a `synchronous` interpretation. Moreover, causally related events in a `result` relation can also be interpreted as synchronous. Thus, as the processing of `synchronous` relations may be confounded by other factors, we will mainly focus on causal relations in the remainder of the paper.

In order to determine how unique this relational distribution is to sentence-initial GFAs, we compare the relational distribution of GFAs to a distribution of inter-sentential (non-GFA) implicit relations. Only when the relational distribution of

Table 6.1: Relational distribution of sentences with a GFA in the dataset.

	Total		Initial		Non-initial	
Relation sense	%	N	%	N	%	N
`result`	24.3	345	28.3	66	23.5	279
`synchronous`	17.3	246	29.2	68	15.0	178
`arg2-as-detail`	16.7	23	17.6	41	16.5	196 7
`arg2-as-manner`	11.6	164	2.6	6	13.3	158
`reason`	10.4	147	2.1	5	12.0	142
`other`	10.2	279	14.6	47	9.4	232

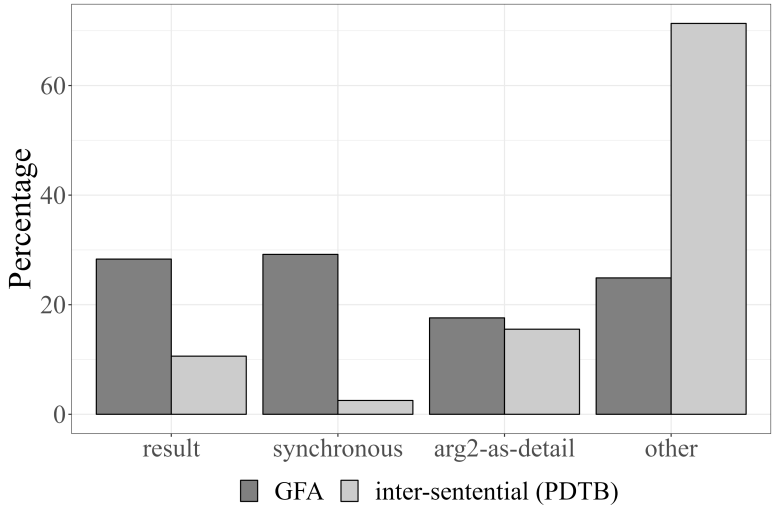

Figure 6.1: Distribution of relation senses for sentence-initial GFAs and inter-sentential implicit relations.

relations containing a GFA is different from that of other clause structures, can the clause structure be a cue for this relational distribution. We therefore extracted the distribution of inter-sentential implicit relations in the PDTB 3, presented in Figure 6.1. The most frequent implicit inter-sentential discourse relation was `conjunction` (24.8% of cases compared to 0.9% of GFAs), followed by `arg2-as-detail` and `reason`. `Result` was the fourth most frequent relation, making up 10.6% of all inter-sentential implicit relations. As can be seen in Figure 6.1, `result` relations are much more frequent in GFAs than in inter-sentential clauses.

6.4.3 Discussion

GFAs frequently occur with `result` relations in natural language, both in non-initial as well as in initial position. More specifically, our corpus investigation has shown that about a quarter of the relations occurring with a GFA are a `result` relation. This is in line with an earlier study showing that GFAs are often translated with an explicit causal connective in Dutch (Hoek & Scholman, 2023), which further supports the hypothesis that GFAs co-occur with causal relations.

Inter-sentential `result` relations in a comparable corpus are much less frequent (about 10%). Readers could use this information in processing discourse relations: `result` relations should be more expected after a GFA. Since our focus is on purely statistical information about the discourse relation based on the clause structure, rather than tense-based semantic cues, we will focus on whether clause structure can serve as a cue for causality, in particular for an upcoming consequence in the following matrix clause.

6.5 Experiment 1: Continuation study

The first step in examining the role of clause structure on discourse relation processing is to investigate whether readers are sensitive to the co-occurrence of GFAs and causal discourse relations. We do so by examining comprehenders' off-line expectations of upcoming discourse relations. If readers are aware of the correlations between clause structure and discourse relations and use this to build expectations of the upcoming discourse structure, readers' expectations should depend on the presence of a GFA. To this end, we conducted a sentence continuation study and examined the relation sense of the continuation that readers provided with the prompt. If clause structure serves as a cue for `result` relations, we should expect to see more `result` continuations in this task in the condition where the prompt is a GFA.

6.5.1 Methodology

Participants

Sixty one native speakers of English (37 female, age range: 18-66 years, mean age: 39) participated in the experiment. They were recruited via Prolific.

Materials

Two versions of twenty-four sentence prompts were created.[4] In the GFA condition, the event was phrased as a gerund free adjunct, as in (50), and in the other condition, it was presented as a full matrix clause (FMC), as in (51). Note that none of the prompts contained verbs that were identified as alternative lexicalizations in the corpus study. The first clause was followed by either a name (in the GFA condition) or a pronoun (in the FMC condition). The start of the matrix clause was provided to ensure that the GFA was not interpreted as the subject of the sentence and to make the insertion of a connective less likely.

(50) Painting his house, Mo ...

(51) Mo was painting his house. He ...

In addition, we created 22 filler items, which contained a causal (n=6), synchronous (n=6) or concessive (n=10) connective. To mirror the target items, half of the fillers started with a subordinate clause and in the other half, the prompt contained a full clause.

Procedure

The experiment was hosted on LingoTurk (Pusse et al., 2016). Participants were instructed to read the prompt and write a logical and grammatical continuation to the sentence in the provided blank. The experimental items were divided across three lists, with each list containing two-third of the experimental items (n=16) and 8 items per condition. The order of the items was randomized for each participant. The study took on average 12 minutes to complete and participants received £1.50 for compensation.

Annotation procedure

The provided continuations were annotated for their relation sense, following the PDTB3 framework. The +SpeechAct and +Belief features were not annotated. Examples of provided continuations to the prompt in Example (51) and their discourse relation annotation can be found below:

(52) Painting his house, Mo ...

[4]This study's design and analysis was preregistered. See: `https://osf.io/bkgf7` (Marchal et al., 2021a)

a. ... created a beautiful mural on his wall. `arg2-as-detail`

b. ... wore overalls to protect his clothes. `result`

c. ... discovered secret text on the wall. `synchronous`

In addition to relation sense, the time frame of the two events was coded as *synchronous* when the two events partly overlap, or *asynchronous* when they do not. Note that this is also a part of the PDTB3 relation sense `synchronous`, but was coded separately here. This was done in order to examine whether the clause structure also raises expectations for how the events in the two clauses are related in time, regardless of the discourse relation. Annotators were blind to the condition, as all prompts were presented to them in the FMC condition. Continuations that were not grammatically correct (n=4) were removed. After training on two batches of 10%, the guidelines were adjusted accordingly and the rest was annotated by two independent annotators.[5] Inter-annotator agreement was calculated on this data set (n=884) and disagreements were resolved by a third annotator, who selected from the relation senses provided by the first two annotators. We allowed for double relation senses if both interpretations were considered likely. The intersection between the two annotators was then taken as the final relation sense. Agreement was calculated using a boot-strapped agreement coefficient on this intersection (Marchal et al., 2022c). This measure calculates agreement on the final overlapping label while taking into account that the expected agreement increases when coders are allowed to provide more than one label. Agreement was moderate: there was an overlapping label for 82% of items ($\kappa = .72$).

Agreement on items (n=434) for which synchronicity was annotated by both annotators was moderate according to κ ($\kappa = .54$ [.42, .64]), but sufficient when considering AC_1 ($AC_1 = .84$ [.79, .89]), mainly due to a strong bias towards synchronicity.[6] All disagreements were resolved by a third annotator.

Statistical analysis

All experimental data in this paper were analyzed using mixed-effects regression analyses, using the `lme4` package in R. We always fit the maximal random effect structure (cf. Barr et al., 2013), reducing it only in case of non-convergence. In these cases, the

[5]Due to low disagreement on the the first two batches, this diverges from the preregistration plan in order to ensure higher data quality.

[6]Kappa underestimates true agreement in scenarios where category prevalence is highly imbalanced by inflating chance agreement, which is corrected for with AC_1 (Hoek & Scholman, 2017).

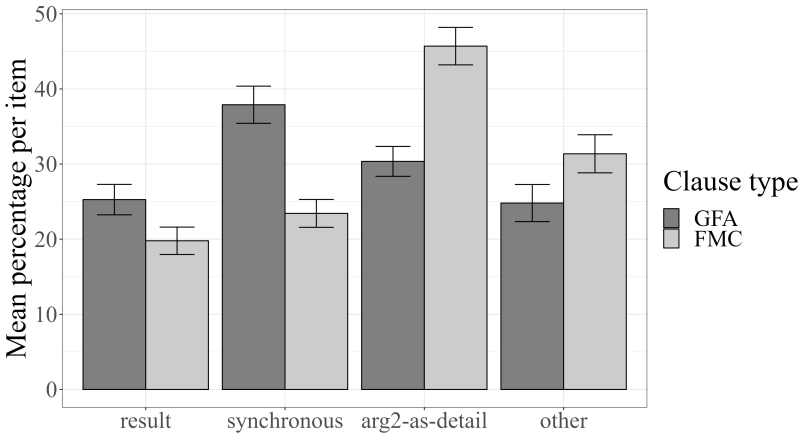

Figure 6.2: Mean by-item percentage of the most frequent relation senses with by-participant standard error bars.

random effect structure will be reported below. Significance of the effects of interest were tested with the package `lmerTest` and condition was sum coded (FMC: -0.5, GFA: 0.5).

6.5.2 Results

The continuations were categorized in twelve different relation senses. All relation senses were identified at least once in each condition. Figure 6.2 shows the proportion of the first relation sense per condition. `arg2-as-detail` is the most frequent relation sense (36.8%) overall, followed by `synchronous` (24.5%). However, as can be seen in Figure 6.2, the distribution of the relation senses differs per condition. Crucially, `result` continuations occurred less frequently in the FMC condition than the GFA condition, which is in line with our hypothesis (β = 0.85, SE = 0.32, z = 2.63, p = 0.009).[7] In 20.0% of the continuations in the GFA condition, a `result` relation had been annotated, compared to 12.3% in the FMC condition.

There were also differences in expectations for other relations between the two conditions. A post-hoc exploratory analysis shows that condition significantly influenced the proportion of `arg2-as-detail` relations, which were more frequent in the FMC condition (β = -0.87, SE = 0.14, z = -5.99, $p < .001$) and `synchronous` relations,

[7]This effect was also significant when only considering the first relation sense (β = 0.65, SE = 0.25, z = 2.62, p = .009). This model only converged after removing the random slope for participant.

which were more frequent in the GFA condition ($\beta = 1.26$, SE = 0.24, z = 5.29, $p < .001$).[8]

Similar to the corpus findings, `synchronous` relations occur often in the GFA condition. To examine whether this is a feature of the tense of the verb, we separately annotated instances that were annotated as a temporal or causal relation (53.9% of all relations) for synchronicity (i.e., whether the events in the two clauses were synchronous or not). Overall, in 80.1% of these instances, the events overlapped significantly in time. However, this effect was stronger in the GFA condition than in the FMC condition ($\beta = 1.53$, SE = 0.55, z = 2.81, $p = .005$), as 90% of continuations in the GFA condition contained synchronous events, compared to 68.8% in the latter. This is in line with the finding that `synchronous` discourse relations are more frequent in the GFA condition, as noted above. However, even within the `result` relations (n = 157), over 90% of the events occur synchronously in both conditions. This suggests that synchronicity is an artefact of the tense of the verb, not of the discourse relation itself. In addition, within the subset of `result` relations, there was no significant difference in event synchronicity between conditions: 95% of events in `result` relations in the FMC condition consisted of synchronous events, compared to 92% in the GFA condition.[9]

6.5.3 Discussion

Our findings show that comprehenders' expectations of upcoming discourse relations are affected by clause structure. The likelihood of various relation senses in the continuations was dependent on the clause type of the prompt. Crucially, `result` relations were more frequent when the prompt was a GFA compared to when it was not. Thus, comprehenders are sensitive to the co-occurrence of the GFA clause structure and `result` discourse relation in natural language, and they use this information to create off-line expectations. Similar to the corpus study, `synchronous` relations were also very frequent in the continuations after a GFA prompt. However, the fact that the vast majority of the events, also in `result` relations, were synchronous, suggests that synchronicity might be a result of the verb tense. The verb tense rather than the clause structure might therefore elicit expectations for `synchronous` relations. Since

[8] P-values for these post-hoc analyses are Bonferroni-corrected.

[9] Due to the small number of observations, this model only converged with a random intercept for items.

verb tense cannot signal causality, the clause structure itself likely functions as a cue for `result` discourse relations.

To sum up, both the corpus investigation as well as the continuation study reveal an effect of clause structure on the distribution of discourse relations. More specifically, `result` relations co-occur more frequently with sentence-initial GFAs than with inter-sentential discourse relations. This is also reflected in readers' off-line expectations for discourse relations, as shown in the present experiment. However, this task only assesses whether readers are sensitive to these statistical correlations, not whether this extends to other levels of discourse processing. If readers use clause structure as a cue for processing discourse relations, they are expected to also incorporate this information during on-line processing and assign meaning to these statistical correlations. In the upcoming studies, we will look at each of these aspects of discourse processing in turn, by examining whether these expectations are also elicited during on-line processing and whether they influence discourse interpretation.

6.6 Experiment 2: Self-paced reading study

The sentence continuation study reveals that the expectations that readers have for discourse relations are different after GFAs than after full matrix clauses. In particular, `result` relations are expected more strongly after GFAs than after FMCs. To investigate whether such expectations also influence on-line processing, we conducted a self-paced reading study. Previous studies have shown that relations that are expected are read faster (Asr & Demberg, 2020; Mak & Sanders, 2013). The results from the continuation study indicate that expectations for `result` relations are stronger after a GFA than after a FMC. We therefore hypothesized that `result` relations are read faster in sentences with a GFA than sentences without.

6.6.1 Methodology

Participants

Eighty native speakers of English living in the United States participated in the study. They were recruited via Prolific. Seven participants failed the attention checks and their data was therefore excluded from the analysis. The age of the remaining 73 participants (46 females) ranged from 19 to 75 years (mean = 33).

Materials

We created event pairs in which the second event was the result of the first event (e.g. painting a house and wearing old clothes).[10] Similar to the sentence continuation study, in the GFA condition (53a), the first event was presented in a gerund free adjunct, and in the FMC condition (53b), it was a full matrix clause. Since a potential slow-down after an FMC might also be caused by the full stop, we also included a control condition. This condition, in which the full matrix clause is connected with the underspecified connective *and* (53c) contains no punctuation at all. The second clause always started with a pronoun to make the clauses more comparable across conditions. The target discourse relation was preceded by an introductory sentence to provide some context. Two additional sentences were added at the end of the text to make them comparable in length to the items of an unrelated study, which was run simultaneously.

(53) *Context:* Mo and his girlfriend had decided to do some renovations before she would move in.

a. Painting the house, [he had been]$_1$ [wearing his old sweater]$_2$ [and ripped jeans]$_3$.

b. Mo was painting the house. [He had been]$_1$ [wearing his old sweater]$_2$ [and ripped jeans]$_3$.

c. Mo was painting the house [and had been]$_1$ [wearing his old sweater]$_2$ [and ripped jeans]$_3$.

Context: His girlfriend thought he looked quite handsome in them. Mo did not feel at ease, because he was used to wearing suits most of the week.

As illustrated above, the second clause was split into three regions for the self-paced reading task. The pre-critical region (1) contained the pronoun/connective and the auxiliary verb phrase. The target region (2) is the region where the `result` relation could be inferred (see *Causality pretest* below) and the spill-over region (3) contains the rest of the sentence.

Causality pretest We pretested the items with 20 native speakers of English in order to ensure that the items were interpreted causally. The pretest consisted of thirty items, as well as fillers, distributed over two lists. Participants were asked to

[10] This study's design and analysis was preregistered. See: `https://osf.io/u3gvs` (Marchal et al., 2022b)

Table 6.2: Mean rating (and standard deviation) of the naturalness of the items per condition.

Condition	M	SD
GFA	5.24	1.58
FMC	5.60	1.61
and	5.66	1.58

first read the context. They were then provided with the two target clauses, ending in the target region. The clauses were presented as full clauses and connected by either *so* or *more specifically*. Participants selected which linking phrase was the most logical. They could also select that none of the connectives fit, if they considered them both to be illogical (but not if they thought it was too formal). Items for which less than 7 out of 10 participants selected the causal connective were improved and retested. Three items were removed after the second pretest, because less than 70% of the participants chose the causal linking phrase.

Naturalness pretest To ensure that the items did not differ in plausibility, the 27 items that were selected based on the causality pretest were also rated for their naturalness. The final spill-over sentence was not included in this pretest, but the spill-over region from the target sentence was. The items were combined with items from an unrelated experiment and fillers (stories that were either incoherent or contained reference or inflection errors) and evenly divided over six lists. 48 native speakers of English, who were based in the US, were asked to rate each text on a 7-point scale, ranging from not natural at all to very natural.

The six items that yielded the largest difference between conditions were removed from further analysis. The final materials thus consisted of 21 experimental items. They differed maximally 1.39 in one condition compared to the grand mean (see Table 6.2), but this was not significant in a linear mixed-effects model.

Procedure

The self-paced reading task was implemented using PCIbex (Schwarz & Zehr, 2021), using a self-paced moving window paradigm. The participants first saw three practice trials, after which they continued to the experimental part, consisting of 75 texts. Participants were randomly assigned to one of the three lists to ensure that each participant saw all items and all three conditions, but only saw every item in one of

the three conditions. The order of the items was randomized for each participant. A quarter of the trials was followed by a comprehension check, a statement to which participants had to respond whether it was true or false. They could take a break halfway during the experiment. In total, the task took about 15-20 minutes to complete and participants received £2.50 for compensation.

Data analysis

Data from participants (n=7) who scored less than 70% on the comprehension questions were excluded from the analysis. In addition, observations from participants on items on which they took longer than a minute (n=5), since this indicates they might have taken a break (overall M = 12.6, SD = 7.7 seconds).[11] Furthermore, reading times above 2500 ms (22 cases) were also removed, as were reading times above or below 2.5 standard deviation away from the participant's mean per region (105 cases). Overall, 4.4% of data points were removed with this outlier analysis. In addition, reading times on one item had to be removed due to a typo in the target region.

We analyze the reading times on the target and the spill-over region separately. To account for multiple comparisons, we applied Holm-Bonferroni corrections. Reading times were log-transformed[12] and condition was treatment-coded, with GFA as the intercept. The models include trial number as covariate and the random effect structure was maximal, with by-participant and by-item random slopes and intercepts for each fixed effect.

6.6.2 Results

The reading times per region are shown in Figure 6.3. In the target region, there is no effect of condition, neither between GFA and FMC (p = 1.00) nor between GFA and the AND condition (p = .29). In the first spill-over region, there is a small numerical difference between reading times in the GFA and the FMC condition, with items in the FMC condition being read faster than in the GFA condition, contrary

[11]This diverges from the preregistration, since the analysis proposed there excludes items from many participants who seem to be 'slow' readers in general.

[12]Unlike our planned analysis, we used log-transformed reading times as the dependent variable due to convergence issues when fitting the model with raw reading times and a Gamma distribution.

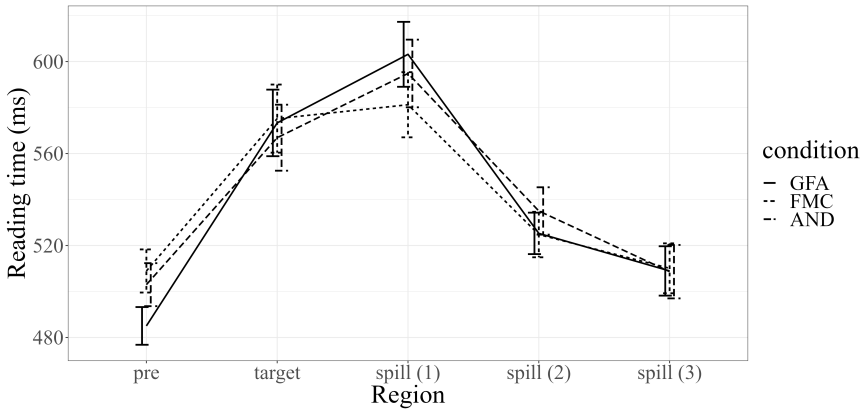

Figure 6.3: Mean reading times per condition per region with standard error bars.

to what was hypothesized. However, this difference is not significant after correcting for multiple comparisons ($p = .06$).[13]

6.6.3 Discussion

Unlike the sentence continuation task, no effect of clause structure on on-line expectations was found in the self-paced reading task. Thus, there is no evidence that statistical correlations between the clause structure and the discourse relation can influence on-line processing. This could be due to the fact that clause structures are not as informative as other types of cues with respect to the discourse relation. Its signal might therefore not be strong enough for readers to incorporate during on-line processing.

It is also possible that no effect was found due to methodological limitations of the self-paced reading task. First of all, differences in the clause structure itself might influence reading times of the subsequent clauses. For example, in the GFA condition, the referential link between the pronoun and its antecedent is longer than in the other two conditions. As a result, the pre-critical region might be harder to process in this condition, as it is harder to resolve the referential pronoun. Such a pronoun-induced slow-down might then show up in the target region, the region where an effect was expected. This would interfere with the hypothesized effect on this region. In addition, at the end of a sentence, participants have to integrate the

[13]A post-hoc analysis revealed that the effect of condition changed over the course of the experiment, such that readers slowed down less over the course of the experiment in the AND condition ($\beta = .002$, SE $= .001$, $t = 2.49$, $p = .01$). Nevertheless, the difference in reading times between the AND and GFA condition was significant neither in the first nor the second half of the experiment.

propositional content of the clause with their knowledge about the the text and their background knowledge. In the FMC condition, these wrap-up effects can be expected to be smaller, since only one clause needs to be integrated, whereas two clauses need to be integrated in the other two conditions. This could explain the trend that the spill-over region was read fastest in the FMC condition. Having found contrasting results in the off-line continuation study and on-line processing study, we now turn to an investigation on the effect of GFAs on the interpretation of discourse relations.

6.7 Experiment 3: Paraphrase selection study

The previous experiments showed that readers' off-line expectations are dependent on clause type in a production task, but that there is no evidence that these expectations also facilitate the processing of expected relations during reading. However, it is possible that readers track the discourse relations that occur with this clause type, but that they are not sensitive enough to this type of information to incorporate it during on-line processing. To assess whether this statistical correlation nevertheless leads to semantic enrichment of the clause type, we investigate the interpretation of discourse relations with and without a GFA. More specifically, we contrast a relation co-occuring more frequently with GFAs (i.e. `result`), with a relation that is relatively more frequent with FMCs: `arg2-as-detail` was found to be more preferred with two matrix clauses in natural language as well as experimentally elicited continuations. Instead of directly examining readers' interpretation of the discourse relation, we assess their preference for a specific clause type based on the discourse relation. If readers associate GFA clauses with `result` more than `arg2-as-detail`, we hypothesize that comprehenders would prefer GFAs to express `result` compared to `arg2-as-detail` relations. In addition, if `arg2-as-detail` relations are more natural with FMC clauses rather than `result` relations, readers should have a preference for this type of structure with `arg2-as-detail` relations.

6.7.1 Methodology

Participants

Forty native speakers (27 female) of English participated in the experiment. They were based in the US and their age ranged from 19 to 69 (mean = 33).

Materials

The materials consisted of eighteen items. For each item, four versions were created.[14] The first clause was structured as either a FMC or a GFA. The event in the first clause was the same across conditions, but in the `result` condition, the first clause was followed by a consequence and in the `arg2-as-detail` condition, the first clause was followed by a specification. The items were pretested to ensure that the target relation was the dominant interpretation. For the `result` relations, this had already been done, as described in Section 6.6.1. For the `arg2-as-detail` relations we followed the same procedure.[15] The discourse relation was preceded by a one-sentence context, which always introduced two characters to license the use of a proper noun at the start of the FMC clause. In addition, for each item in each condition, two paraphrases were created: one in which the first clause was presented as a gerund free adjunct, and one in which it was a full matrix clause. An example of an item in the two conditions is presented below.

(54) `result`

a. Mo was painting the house. He wore an old sweater. FMC

b. Painting the house, Mo wore an old sweater. GFA

(55) `detail`

a. Mo was painting the house. He painted the walls blue. FMC

b. Painting the house, he painted the walls blue. GFA

Procedure

Participants were asked to select which of the two paraphrases (the GFA or the FMC) sounded the most natural. The items were distributed over two lists, so that each item was presented only in one of the conditions to the participants. The order of item presentation was automatically randomized per participant. Attention checks consisted of 4 additional items in which participants were instructed in the prompt which paraphrase to select. Other than that, the experiment did not contain any fillers. The task took about 7 minutes to complete and participants received £0.90 for compensation.

[14]This study's design and analysis was preregistered. See: `https://osf.io/cmf3s` (Marchal et al., 2022a)

[15]The cut-off point was lowered to 65%.

Table 6.3: Proportion of selected clause types per condition.

relation sense	FMC	GFA
`arg2-as-detail`	0.52	0.48
`result`	0.54	0.46

Analysis

Data from four participants who did not pass the attention check was removed and replaced with data from new participants. We used a binomial mixed-effects logistic regression analysis to examine whether the option with the GFA was selected more often in the `result` condition compared to the FMC condition. The model included condition as a fixed effect (`result`: -1, `arg2-as-detail`: 1) and by-participant and by-item random slopes and intercepts.

6.7.2 Results

Table 6.3 displays the proportion of selections per condition. There was no significant difference in participants' preference for a specific verb clause in one of the two conditions.

In each item and in each condition, both options had been selected at least once. This not only shows that the preference for a specific combination of verb type and discourse relation is not deterministic, but also that both options were considered plausible in both conditions. In addition, there was considerable variation between participants in their overall preference for one of the two clause types. With the exception of one participant who selected GFAs exclusively, all participants varied in their preference for GFAs vs. FMCs between items. This indicates that a reader's preference for one of the two clause types depends on external factors.

6.7.3 Discussion

In this paraphrase selection study, readers did not show a preference for `result` relations in GFAs. It could be noted that in our corpus study, `detail` relations comprise only a slightly larger proportion of relations in the inter-sentential discourse relations compared to the GFAs. Nevertheless, this is not the case for `result` relations, which we had hypothesized to drive the effect. In addition, in the sentence continuation study, which contained sentence structures identical to the ones here, there was a

strong preference for `arg2-as-detail` in the FMC compared to the GFA. This suggests that `result` relations should be more natural with GFAs and `arg2-as-detail` with FMCs. However this was not found in the present study. Possibly, readers are aware of the statistical correlations between discourse relations and clause structure – as evidenced in the first study –, but does this statistical information not influence discourse processing and representation – as suggested by the lack of an effect in the second and third study.

One limitation of the present experiment is that participants were forced to select one of the two options and we therefore do not know whether the other option was also viable. We chose this approach because the preference for one of the two options was hypothesized to be relatively small and we thought that participants might have been undecided when asked to rate the items on a scale (i.e. rated both items with the same score). However, allowing participants to provide their (relative) preference for one of the items (i.e. by allowing them to assign different weights to each option) might have allowed us to pick up on more fine-grained effects. This might have been especially insightful for items for which there was a strong preference for one of the clause types, possibly due to the context, or for participants who selected one of the clause types in the vast majority of cases. We could then expect the preference of these participants or workers for a clause type to be less strong in the `result` condition compared to the `arg2-as-detail` condition.

6.8 General discussion

The aim of the present research was to investigate the influence of statistical correlations between a linguistic phenomenon and discourse structure on discourse-level processing. We approached this research question by examining the relation between a specific clause type, GFAs, and discourse relations. Corpus distributions showed that in natural language, GFAs occur frequently with causal and synchronous discourse relations. Sentence-initial GFAs, in particular, frequently co-occurred with `result` relations. We hypothesized that comprehenders are sensitive to these correlations and that they would affect discourse processing at different levels. More specifically, we examined off-line production (do these correlations elicit expectations?), on-line reading (do these correlations facilitate the processing of expected relations?) and off-line interpretations (do these correlations influence comprehenders' biases?) to gain a comprehensive overview of the role of statistical co-occurrence on language processing.

The finding that the probability of a discourse relation depends on the clause structure was replicated in a production task, where participants had to provide continuations to short prompts either consisting of a GFA and a main clause, or of two main clauses. This task revealed that readers have stronger expectations for `result` and `synchronous` relations in the GFA condition, but for `arg2-as-detail` relations after a full matrix clause. However, contrary to our hypothesis, this expectation for a `result` relation after a GFA was not found to facilitate the processing of these relations in a self-paced reading task. There were no differences in reading times between `result` relations following a GFA, compared to following an FMC. In addition, we did not find any evidence that the co-occurrence between clause type and specific discourse relations in production influences discourse representation. Readers did not have a preference for `result` relations with a GFA and `arg2-as-detail` relations with an FMC, despite the fact that these relations co-occur with these clause structures in natural language. In the following paragraphs we will provide possible explanations for these results and discuss the implications of our findings.

6.8.1 Features of the signal

In Chapter 3, we discussed five features that could influence whether a signal has an effect on processing and representation: functionality, informativity, immutability and lexicality. These features might also explain the limited effect of gerund free adjuncts. Here, we will discuss how these features relate to gerund free adjuncts.[16]

First, clause structure seems to co-occur with a wide variety of relation senses: twenty different relation senses were identified with GFAs in the corpus and twelve in the continuation study. This is not only true of clause structure, but of all non-connective signals for discourse relations. After all, their primary function is not to signal the relation (cf. *functionality* feature, Chapter 3, Section 3.3.4), but to convey propositional meaning (e.g., polarity in the case of negation). Furthermore, discourse connectives often signal only one or a few relations. Due to the higher entropy distribution, non-lexical cues provide much less information (cf. *informativity* feature, Chapter 3). As argued in Chapter 3 (Section 3.3.4), they might therefore be less beneficial for readers to take into account when processing language. As a result, their statistical correlation might not lead to semantic enrichment. Although

[16]We do not consider *immutability* as an explanation for the lack of an effect, as there is evidence from other studies that discourse cues that are context-dependent influence representation and processing (cf. Crible & Demberg, 2020; Kehler et al., 2008; Sanders & Noordman, 2000; Rohde & Horton, 2014, see Chapter 3, Section 3.6).

we define discourse relation signals as any cue that provides any information about the discourse relation in Chapter 3 (Section 3.2), there might be a threshold for the amount of information that the signal provides before comprehenders will use it.

Another explanation for the lack of an effect in most of the experiments presented in this chapter might have to do with the nature of the cue. Possibly, statistical correlations between a cue and a relation alone does not suffice for a meaning enrichment of that cue. Previous studies investigating non-connective cues often focused on a semantic relation between the cue and the relation that it signaled. For example, negation has a conceptual link with `concession`. Following the taxonomy proposed by Hoek et al. (2019b), these cues are in agreement with the discourse relation (i.e. *agreement* feature). Clause structure, on the other hand, does not provide any semantic information about a causal relation between the segments. Thus, the only possibility for readers to associate GFAs with specific relation senses is through statistical correlations. We did not find any evidence that readers are able to use this information in the processing and interpretation of discourse relations. Conceptual links between form and function (i.e. agreement) might be necessary to facilitate discourse-level processing, at least in the absence of strong lexical cues. Specifically, implicit causality verbs are also not in agreement with the discourse relation meaning, but still strongly influence discourse expectations in both off-line and on-line processing (Kehler et al., 2008; Rohde & Horton, 2014; Hoek et al., 2021a). Although Crible & Pickering (2020) do find evidence that a syntactic cue can facilitate the processing of discourse relations it signals, this effect only shows up when participants are actively disambiguating the relation. Further research is necessary to examine task-related differences in the effect of discourse relation signals. Here, we focused on readers' sensitivity to such a cue during natural reading.

A further difference between GFAs and connectives as signals of discourse relations is that connectives are lexical. This could also explain why no effect was found in the present study. It is unclear whether signals need to be lexical in order to facilitate online processing. Previous studies investigating non-lexical cues show mixed findings on the effect of these cues for processing (cf. Grisot & Blochowiak, 2017, for tense and cf. Grisot & Blochowiak, 2017, for parallelism). With respect to representation, there is evidence that non-lexical cues influence the interpretation of discourse relations (see Rohde et al., 2017 for complementizers and Marx et al., 2024 for eventuality). In the present study, we find no effect of clause structure on neither processing nor representation.

Finally, certain connectives and discourse relations are notoriously difficult to acquire. Even adults do not perform at ceiling when their connective comprehension is tested, especially with low frequent connectives (Zufferey & Gygax, 2020b; Tskhovrebova et al., 2022a; Scholman et al., 2024a). Importantly, performance on such tasks is correlated with linguistic experience (Zufferey & Gygax, 2020b; Scholman et al., 2024a), as is readers' sensitivity to contextual discourse cues (Scholman et al., 2020). Since the discourse cue examined here is even more subtle than a lexical cue and also rather infrequent, even more exposure might be needed to infer meaning from the co-occurrence of the signal and the discourse relation. It is possible that effects would only have been found for highly literate readers. We leave this issue for further research.

6.8.2 Converging evidence: Mixed findings

The present study aimed to provide converging evidence on the role of gerund free adjuncts as discourse signals (cf. Chapter 2, Section 2.4.3). However, the different studies showed mixed findings: although we found evidence for the informativity of gerund free adjuncts in the corpus studies, no effects of this clause type was found on the processing and representation in Experiment 2 and 3. One question that remains is why clause structure did affect the continuations that participants provided in Experiment 1, in line with the results from the corpus study. The results of the first experiment suggests that readers are sensitive to the co-occurrence between clause type and discourse relations to some extent. Note that the nature of this task is very different from the other experiments. The first study differed from the other studies in two respects, relating to the off-line and on-line nature and the production versus comprehension aspect. Regarding the first difference, the sentence continuation study consisted of an off-line task, in which participants were not limited in the amount of time they needed to complete the sentence. The self-paced reading task, on the other hand, measured on-line processing. The demands in such an on-line task are much higher, as participants are not only engaged in predicting upcoming material, but also with parsing the syntactic structure, retrieving the meaning of words and propositions, and integrating the clauses. Readers may only be sensitive to GFAs as a cue for `result` relations when they have resources available to attend to such a cue. It should be noted that the paraphrase selection study also measured off-line processes, but no effect was found in this experiment. However, this task did not assess discourse expectations directly. If it is indeed the case that the effect of GFA on on-line processing was mitigated by the task demands, it is possible that

readers with strong expectations for causality after a GFA are still sensitive to the effect. It would be interesting to see if readers who show stronger expectations for causality after GFAs in a production task, are also more sensitive to this cue in on-line processing.[17]

A second difference between the sentence continuation task and the two other experiments is that the first assessed production, whereas the other two experiments focused on comprehension. The production task did not restrict the content of the upcoming clause in any way, whereas the upcoming clause was provided to participants in the two other tasks. It is possible that we might not have been aware of an additional characteristic of the sentences with GFAs in the corpus and production study that determines the processing of gerund free adjunct. In such a scenario, the items constructed for the follow-up experiments might not have contained these features. Using more naturalistic texts could solve this issue, but such an approach also has its drawbacks, as it makes it more difficult to control for factors which are known to influence reading times (e.g., length, frequency, required background knowledge). Another explanation for why the effect only shows up in production and not in the comprehension tasks could be that readers are aware of the correlations (e.g., as a convention), but do not associate any meaning with it. In that case, we would expect to see an effect of GFA in a production task, as readers adhere to patterns that they know, but not when assessing discourse representations. This explanation is consistent with our findings.

6.8.3 Limitations

There are several limitations to this study. First of all, even though we do not find evidence in favor of the hypothesis that readers use gerund free adjuncts in the processing of `result`, this does not mean that there is no effect. Absence of evidence is not evidence of absence. Possibly, the effect of gerund free adjuncts is too small to be detected with the number of participants and items in our study. Note, however, that in Chapters 4 and 5, we did find evidence for the facilitating effect of a connective using a similar methodology and number of participants. This suggests that the effect of gerund free adjuncts is smaller compared to connectives, but this should be tested further. Ideally this should be tested in an experiment manipulating the presence

[17]Note that the participants in the first experiment were sampled from the same population as participants in the last two experiments.

of both types of cues, yielding a direct comparison on the effect of connective and non-connective cues. We leave this for future work.

Another limitation of the present study is that gerund free adjuncts also often co-occurred with `synchronous` interpretations. Rather than mapping this cue onto `result` relations, readers may associate this cue with `synchronous` relations. We controlled for this by ensuring that the events in all materials in Experiment 2 and 3 could also occur synchronously. Thus, even when readers associate gerund free adjuncts with `synchronous` relations, we should still expect to see a facilitation or preference for gerund free adjuncts in the relations in our study. However, it is unclear how readers process multi-interpretable relations (cf. Rohde et al., 2016; Scholman, 2019) and how they interpret linguistic signals for this type of relations. To illustrate, readers might not have had the relation interpretation that they associate with the cue. Further research on multi-interpretable relations could shed more light on how relations with multiple interpretations are processed. For example, are readers aware that multiple interpretations hold? If so, are such relations more difficult to process? And how does this affect readers' sensitivity to linguistic signals for these relations? These questions could be examined in further research.

6.9 Conclusion

The present study examined whether readers' sensitivity to linguistic signals for discourse relations is dependent on characteristics of the signal (Research Goal 1, Chapter 1). Specifically, we extended previous work on the role of connectives in discourse processing (e.g. Millis & Just, 1994; Cozijn et al., 2011; Van Silfhout et al., 2015, see also Chapters 4 and 5) to examining the effect of a non-lexical cue that is not specialized for signaling the discourse relation (cf. features presented in Chapter 3). The present study investigated whether statistical information, namely the co-occurrence of GFAs and causal discourse relations in natural language, influences readers' discourse processing. We found evidence for this in a production task, showing that readers' off-line expectations for discourse relations depends on the clause type. However, this effect did not show up in experiments examining on-line processing and discourse interpretation. The effect of statistical correlations on discourse processing thus seems to be limited. Possibly, a semantic relation between the linguistic cue and its meaning (i.e. agreement) is necessary for it to influence higher-level discourse processing. Further research is necessary to gain more insight in the exact conditions for which statistical correlations influence which aspects of discourse processing, especially with

respect to the informativity, lexicality and agreement of the signal. We will return to this issue in Chapter 8. First, we focus on another source of information that could be relevant in inferring discourse relations: domain knowledge.

Data availability

All materials, data and analysis code of all experiments reported in this study have been made publicly available at: `https://osf.io/heqsu/`.

Funding

This work was supported by the German Research Foundation (DFG) under Grant 232722074 - SFB 1102 ("Information Density and Linguistic Encoding").

Dissemination

This chapter is based on the following journal article:

- **Marchal, M.**, Scholman, M.C.J., & Demberg, V. (2023). How statistical correlations influence discourse-level processing: Clause type as a cue for discourse relations. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *50*(5), 796-807.

In addition, (parts of) this research has been presented at the following conferences:

- **Marchal, M.**, Scholman, M.C.J., & Demberg, V. (2022). Clause structure as an alternative cue for discourse relations. *Talk at the 32nd Annual Meeting of the Society for Text and Discourse, online, 19-21 July.*
- **Marchal, M.**, Scholman, M.C.J., & Demberg, V. (2022). How statistical correlations influence discourse-level processing: Clause type as a cue for discourse relations. *Talk at DisCorX, Bern, Switzerland, 16-17 November.*

Chapter 7

Beyond language
Domain knowledge as a cue for inferring discourse relations

The previous chapters have focused on the role of linguistic signals for discourse relations, but, as discussed in Chapter 2, readers can also rely on extra-linguistic sources of information. It is generally assumed that readers draw on their background knowledge to make inferences about information that is left implicit in the text. However, readers may differ in how much background knowledge they have, which may impact their text understanding. The study presented in this chapter investigates the role of domain knowledge in discourse relation interpretation, in order to examine how readers with high vs. low domain knowledge differ in their discourse relation inferences. We compare interpretations of experts from the field of economics and biomedical sciences in scientific biomedical texts as well as more easily accessible economic texts. The results show that high-knowledge readers from the biomedical domain are better at inferring the correct relation interpretation in biomedical texts compared to low-knowledge readers, but such an effect was not found for the economic domain. The results also suggest that, in the absence of domain knowledge, readers exploit linguistic signals other than connectives to infer the discourse relation, but domain knowledge is sometimes required to exploit these cues. The study provides insight into the impact of domain knowledge on discourse relation inferencing and how readers interpret discourse relations when they lack the required domain knowledge.

It also shows that how readers exploit linguistic signals depends on characteristics of the reader.[1]

[1]This chapter is based on (and in parts identical to) the following publication: **Marchal, M.**, Scholman, M.C.J., & Demberg, V. (2022). The effect of domain knowledge on discourse relation inferences: Relation marking and interpretation strategies. *Dialogue & Discourse*, 13(2), 49-78. doi:10.5210/dad.2022.202

7.1 Introduction

Successful text comprehension requires readers to understand how the various concepts in a text are related and to integrate the text with background knowledge already available to the reader (see Van den Broek, 2010). One part of text comprehension where background knowledge could help is establishing the connections between information in different parts of the text. Prior studies have suggested that background knowledge indeed supports the inference of discourse relations, assuming that this knowledge is activated to fill in information that is missing in the text (Noordman & Vonk, 2015), but the role of domain knowledge in inferring discourse relations is still unclear. Earlier work has often focused on the role of background knowledge on text comprehension or recall (for an overview, see Smith et al., 2021), but how discourse relation inferences differ between high- and low-knowledge readers has not been investigated systematically.

Moreover, if the reader cannot draw on background knowledge to infer the relation, they might resort to other strategies. What other factors guide the interpretation of discourse relations for low-knowledge compared to high-knowledge readers is not yet known. Here, we examine three possible strategies. Firstly, as previously discussed, linguistic cues play a crucial role in relational inferences. Not only connectives, but also non-connective cues can signal discourse relations. However, these cues are ambiguous and might need to be supplemented with non-textual information (e.g. background knowledge) to infer the intended relation. Secondly, cognitive biases in processing discourse relations, discussed in 2 (see also Chapter 4), might influence relation interpretation in the absence of domain knowledge. Thirdly, readers might resort to more shallow processing when they are not able to infer the discourse relation. This will be discussed in Section 7.2.2.

The goal of the study presented here is therefore two-fold. First, it aims to investigate whether domain knowledge leads to more correct interpretations of discourse relations in the absence of a connective. This will be assessed by eliciting discourse relation interpretations from high- and low-knowledge readers and comparing them to a gold label annotation. Second, this research sets out to explore how readers infer the discourse relation if they lack the necessary domain knowledge.

In the next section, we will first review previous research on the role of domain knowledge in discourse inferences and discuss which factors influence discourse relation interpretation and could help low-knowledge readers to infer discourse relations for which domain knowledge is required. The hypotheses are outlined in Section 7.3,

followed by a description of the methodology. The results are presented in Section 7.5. These are subsequently discussed in the final section.

7.2 Background

7.2.1 The role of domain knowledge in discourse inferences

Several models of language comprehension suggest that readers exploit their knowledge base about the concepts in the text to create a coherent representation of the text (e.g. *Construction-Integration model*, Kintsch & Van Dijk, 1978, *Landscape Model*; Van den Broek, 2010). This knowledge is activated when reading about relevant concepts in the texts, after which the information is retrieved from the long-term memory and can then be integrated with the representation that has been made of the text so far. In addition, reading about these concepts activates additional relevant information in the knowledge base, which can in turn influence predictions about subsequent text (cf. Venhuizen et al., 2019; Ferreira & Chantavarin, 2018). For example, comprehenders adopt general world knowledge in a similar way to linguistic cues to predict event structures within sentences (Milburn et al., 2016). Similarly, readers have been shown to use general world knowledge to make predictions beyond the sentence boundary. For example, Kuperberg et al. (2011) show that events that are highly causally related to events presented in preceding sentences based on world knowledge evoke a smaller N400 than events that have a less strong causal relationship with the preceding context. Furthermore, such predictions are influenced by the presence of connectives (cf. Xiang & Kuperberg, 2015; Köhne-Fuetterer et al., 2021). Reading comprehension is thus a dynamic process in which bottom-up and top-down processes are combined. If a discourse relation is not expressed linguistically, readers can utilize information from the knowledge base on how the events in the text are related to establish coherence.

There are different types of non-linguistic knowledge that readers can have. In the literature, a distinction is sometimes made between background knowledge (i.e. all the knowledge the reader can bring to the text), world knowledge, and domain knowledge (e.g., Smith et al., 2021). Domain knowledge is a type of background knowledge about a specific area (e.g. *apoptosis* is natural cell death). In this sense, it could be distinguished from general world knowledge (e.g. the sky is blue), which is considered to be available to almost every reader. It should be noted that we do not assume that if a reader has domain knowledge about the topic of the text, they will know

how all concepts in the text are related, nor do we assume that this knowledge is required to infer every relation. Some concepts in the text might still be unknown to high-knowledge readers, and some textual relations can also be inferred without knowledge about the domain of the text. However, we do hypothesize that readers with domain knowledge might find it easier to infer the discourse relations in a text from their domain of expertise compared to readers without this specific knowledge, because they are more familiar with the concepts discussed in the text and can rely on an already existing knowledge structure.

Empirical evidence that readers benefit from domain knowledge in making discourse inferences comes from various studies on the influence of coherence marking on reading comprehension. This line of research has repeatedly shown a 'reverse cohesion' effect: in general, low-knowledge readers benefit from texts with high coherence marking, whereas high-knowledge readers show better comprehension after reading a low-cohesive text (McNamara & Kintsch, 1996; O'Reilly & McNamara, 2007; Kamalski et al., 2008; McNamara, 2001). Linguistic marking of coherence enables low-knowledge readers to understand how the concepts in a text are related. In the absence of such cues, comprehension will be impaired. For high-knowledge readers, on the other hand, a text with low cohesion induces them to employ their knowledge base to fill in the gaps in the text. Connecting the concepts from the text with those in their long-term memory then leads to deeper comprehension (McNamara & Kintsch, 1996). These studies have focused on the role of domain knowledge in text comprehension in general, but do not reveal how this influences the interpretation of discourse relations. Examining how low- and high-knowledge readers interpret discourse relations differently can provide more insights into the qualitative differences in text comprehension for these groups of readers. In addition, little is known about strategies that low-knowledge readers may have to comprehend an out-of-domain text. This will be addressed in the current study.

7.2.2 Strategies for inferring discourse relations

In addition to discourse connectives and background knowledge, several other factors have been suggested to influence discourse inferences. In cases where readers lack the domain knowledge to infer the discourse relation, and no connective is available to signal the relation, readers might resort to other strategies to establish coherence. Here, I explore three strategies. Specifically, readers might (i) use non-connective linguistic signals for coherence relations, (ii) rely on cognitive biases for relational inferences, or (iii) process the text more shallowly. How these factors influence discourse relation

inferences and how they might guide the interpretations of low-knowledge readers is outlined in more detail below.

Exploiting discourse relational cues

Non-connective linguistic cues for discourse relations are omnipresent. Das & Taboada (2018a) found that more than 80% of the relations in the RST Signalling Corpus were signaled by means of cues other than connectives. Non-connective linguistic cues for discourse relations might thus well play a role in inferring discourse relations in the absence of the required background knowledge as well.

However, non-connective cues are not as evident cues of discourse relations as connectives and readers might therefore not always adopt them. As discussed in Chapter 3, non-connective cues differ from discourse connectives in that their main function is to convey propositional content, instead of signaling the discourse relation. Furthermore, contrary to connectives, non-connective cues need not be lexical and their form is often context-dependent. Readers might thus not always be aware that they are also a cue for the discourse relation. Furthermore, these elements are more ambiguous than discourse connectives in that they can correlate with a large variety of discourse relations. In other words, they are less informative. Thus, even though such patterns provide cues about the discourse relation, readers might not pick up on which exact relation they signal.

Still, there is some prior literature showing that readers are sensitive to cues other than connectives in processing discourse relations (see Chapter 3 for a more elaborate discussion). For example, readers are better able to infer `contrast` relations when the segments have parallel syntactic structure or contain antonyms (Crible, 2020). In addition, quantifiers (e.g. *several*) in the context have been shown to elicit expectations about upcoming list relations (Scholman et al., 2020; Tskhovrebova et al., 2022a). However, this effect depends on the reader's linguistic experience (Scholman et al., 2020). This suggests that not all readers pick up on discourse relational cues equally well.

Interestingly, corpus research has shown that the distribution of such cues seems to be different in explicit and implicit relations (cf. Sporleder & Lascarides, 2008). More specifically, non-connective linguistic signals for discourse relations appear to be more frequent in implicit relations than in explicit relations (Hoek et al., 2019b). In addition, Crible (2020) found that non-connective cues are more likely to occur in explicit relations if the connective is ambiguous. These findings are in line with Gricean's maxim of quantity to not make a message more informative than necessary

(Grice, 1975) (see Asr & Demberg, 2012, 2015, for an explanation based on the Uniform Information Density Hypothesis).

The present study aims to investigate how readers' sensitivity to discourse relational cues interacts with their domain knowledge. To manipulate the degree to which linguistic signals might be present in the text, both originally implicit and originally explicit relations will be used in the current study.[2] For the originally explicit relations, we remove the connective to create implicit versions. We call these instances in which the original connective has been removed *implicitated* relations (following e.g. Hoek et al., 2017, on implicitation in translation). Previous studies on discourse relation inferences have used either implicit (Scholman & Demberg, 2017a; Yung et al., 2019) or implicitated relations (Sanders et al., 1992), but have not compared readers' accuracy on the two types of relations. Since relations that are implicit likely contain more discourse relational cues than relations that have been implicitated, we expect readers to be better at inferring implicit relations than implicitated relations. We thus predict that agreement on implicitated relations will be lower than on originally implicit relations. Note that implicit relations might also be left unmarked because they are easy to infer based on general world knowledge and will therefore yield higher accuracy. Likewise, implicitated relations might be more difficult because the meaning changes when the connective is removed. We will return to this issue in the discussion.

Crucially, we also predict a possible interaction with domain knowledge here. If the text contains relational cues, both high- and low-knowledge readers might be able to employ these to infer the discourse relation. The effect of domain knowledge will then be moderated. However, if the amount of linguistic information for the discourse relation is limited, as in the case of implicitated relations, high-knowledge readers can still rely on their domain knowledge to infer the relation. Low-knowledge readers, on the other hand, do not have this information at their disposal and will then struggle with inferring the intended relation, leading to lower agreement. The effect of domain knowledge is therefore hypothesized to be larger for implicitated than for implicit relations.

[2]Manipulating the materials by adding or removing these cues was not deemed suitable for the present study, given the relatively limited insights that are currently available regarding the variety of non-connective relational cues and their effects. Furthermore, readers with different levels of domain knowledge might make use of different types of signals, but we do not know beforehand what these signals might be. We therefore use natural text to be able to explore what such cues might be in a qualitative analysis of the results.

Cognitive biases in relation inferences

Another way in which readers might infer discourse relations in the absence of other information is by relying on cognitive biases towards certain discourse relation inferences. According to the *continuity hypothesis* (Segal et al., 1991; Murray, 1997), readers prefer to interpret information in a text as being temporally and causally continuous. More specifically, it suggests that readers tend to relate sentences in an additive, temporal or causal way. Similarly, the *causality-by-default hypothesis* (Sanders, 2005) states that readers have a bias to infer causal relations between the segments in a text. Several corpus-based and experimental studies have provided evidence for these hypotheses (see Chapter 2).

With respect to discourse inference strategies in cases where domain knowledge is required to interpret the relation correctly, it can then be hypothesized that readers will default to inferring a causal or another type of continuous discourse relation if no other relation can be inferred. Only in cases where readers have the necessary background knowledge to infer the relation, they might not rely on these cognitive biases and infer a less-expected discourse relation.

The hypothesis that readers infer expected relations is also supported by a shallow processing account for discourse relations in the absence of domain knowledge. According to Graesser et al. (1994), readers with less background knowledge process text less deeply and might even abandon their search for coherence. Several experimental studies have shown that readers are indeed less likely to make inferences during reading when they lack background knowledge (Noordman et al., 2015, 1992). This suggests that low-knowledge readers process discourse relations more shallowly. Scholman (2019) shows that shallow processing might lead to a higher susceptibility for cognitive biases in relation interpretation. In her study, readers interpret `instantiation` and `specification` relations less often as being argumentative when being forced to process the relation more deeply (i.e. by first summarizing the text). Thus, according to a shallow-processing account, low-knowledge readers who process the text more shallowly might therefore have a stronger preference for continuous and causal discourse relations.

Underspecified interpretations

Finally, low-knowledge readers could abstain from committing to a specific discourse relation, but rather make an approximate assumption about the meaning. For example, readers might infer that there is a negative, or adversative, relation between the

segments in (56), but not determine whether it is a `concession` (i.e., one of the segments raises an expectation that is denied in the other segment) or a `contrast` (i.e., the two segments present two different concepts). In the `concession` reading, Juan knew that his girlfriend would be satisfied with just a drink, but ordered much more despite that. In the `contrast` interpretation, Juan's extensive order is compared to the small order of his girlfriend.

(56) Juan ordered everything on the menu. His girlfriend only wanted something to drink.

Such underspecified interpretations can arise from two causes. On the one hand, it might be a result of shallow processing, similar to a preference for cognitively expected relations. If readers process a text shallowly, they might be satisfied with only inferring that the relation is negative and not wish to specify it further, as this would require more effort. On the other hand, such underspecified interpretations might also arise from uncertainty about the discourse relation. Even if a low-knowledge reader processes the text deeply, they might still remain uncertain about the specific relation sense when they lack the required domain knowledge. For example, readers might not be able to determine whether the relation in (56) above, is a `concession` or a `contrast` relation, despite wishing to do so. If they are nevertheless able to infer some features of the discourse relation (e.g. that it is a negative relation), they might still infer such an underspecified relation, rather than committing to a specific relation that could be wrong.

Participants can express uncertainty about the relation sense through their connectives. For example, the connective 'but' is underspecified regarding its relational sense: it can be used to express both `contrast` and `concession` (Asr & Demberg, 2020). A connective like *by contrast*, on the other hand, is more specific, as it can only be used in `contrast` relations. Similarly, *nevertheless* specifies the relation for `concession`. Readers who retain underspecified interpretations might therefore prefer to provide ambiguous connectives, such as *but*. If readers make specific relation interpretations, they will insert more specific connectives, like *nevertheless*.

7.3 The present study

Background knowledge has been assumed to help readers to infer discourse relations, but it is still unclear how discourse relation inferences differ between high- and low-knowledge readers, both with respect to the quality of the inference as well as the

cues that these different types of readers use. In this study, we manipulate domain knowledge by presenting experts from economics and biomedical sciences with texts that either stem from their domain of expertise (e.g. biomedical research papers in the case of biomedical experts) or from the other domain (e.g. biomedical research papers for economists). The biomedical texts included in this study stem from research papers, which were written for experts in the field. These texts are likely difficult to understand for readers without domain knowledge. The economic texts stem from newspaper articles, which were written for a broader audience. The effect of domain knowledge may therefore be less strong in this genre. Nevertheless, since the topic of the newspaper texts focuses on a specific domain, these texts may still be easier to understand for readers who are familiar with that domain than those who are not.

Discourse interpretations were elicited using a connective insertion task (Yung et al., 2019) and compared to gold label annotations. To examine the use of non-connective linguistic signals for discourse relations (see Section 2.2.1), the relations were either originally implicit or implicitated for the purposes of the current study.

The first research question that this study will address is:

- Do high-knowledge readers make more accurate discourse inferences than low-knowledge readers?

If high-knowledge readers employ their knowledge base when inferring how segments in a text are related (cf. Noordman & Vonk, 2015), high-knowledge readers are expected to infer the relation correctly more often than low-knowledge readers.

Secondly, when required to make an inference about a discourse relation, low-knowledge readers might take several approaches to establish coherence in the text. The second aim of the study is therefore to investigate:

- What inferences do readers make if their domain knowledge is insufficient to infer the discourse relation?

Based on the discussion above, we can formulate three hypotheses about what readers will do in the absence of domain knowledge:

(a) Readers use non-connective linguistic signals to infer the discourse relation.

(b) Readers resort to default interpretation strategies based on cognitive biases for continuity and causality.

(c) Readers make less precise interpretations about the discourse relation.

Since implicit relations have been suggested to contain more non-connective linguistic signals than implicitated relations, we hypothesize that these relations will be easier to infer. Moreover, we predict that this effect is stronger for low-knowledge readers, as they are hypothesized to be unable to compensate for their lack of domain knowledge in implicitated relations. Note that it might also be the case that relations are left implicit for other reasons, for example because they are easier to infer on the basis of general world knowledge. We will therefore also examine qualitative differences in the presence of linguistic signals in items on which high- and low-knowledge readers differ. In addition, if low-knowledge readers' interpretations are guided by their cognitive biases, it is predicted that these readers will infer more continuous and causal discourse relations than high-knowledge readers. If the third hypothesis is true, low-knowledge participants are predicted to insert more ambiguous connectives than high-knowledge readers, as they reflect their underspecified interpretation better than specific connectives.

These hypotheses are not mutually exclusive. Readers might attempt to use linguistic correlates for discourse relations to make the inference, but leave the relation underspecified if these cues are not sufficient to make a precise inference. Similarly, if these discourse relational signals are absent, readers may interpret the relation as being continuous, but not further commit their interpretation to a specific type of continuous relation.

7.4 Methodology

7.4.1 Participants

We recruited students and graduates in the field of economics and biomedical sciences on Prolific for a prescreening study. More specifically, five hundred workers participated that had registered on Prolific that their subject of study was in the field of economics or biomedical sciences.[3] The prescreening study served to further ensure expertise in one of the two domains and assess familiarity with each domain. Participants were presented with a short questionnaire assessing their demographic background and familiarity with both fields. The latter involved questions about their study and work experience in the field. In addition, we assessed participants' knowl-

[3]When registering on Prolific, participants had indicated that their field of study was either in economics, accounting and/or finance or in biomedical sciences, genetics, biology, biological sciences and/or biochemistry.

edge by asking them to indicate their familiarity with 10 concepts that are specific to the two domains extracted from the two corpora (e.g. *volatility*, *phosphorylation*). For each of these concepts, participants could indicate that they either did not know the term (which was coded as 1), had heard of the term while not being able to describe it (coded as 2), or would also be able to provide a description of the term (coded as 3). Five filler concepts for each domain consisting of terms that were deemed familiar for non-experts as well (e.g. *interest, DNA*) were included as an attention check. Despite relying on participant's self-assessment, this method has been shown to correlate well with participants' performance on textbook questions (Mehti, 2024) and can therefore be deemed reliable.

In order to ensure that our study participants were knowledgeable in their own field of expertise, but not in the other field, we selected only those participants that met the following criteria for the final experiment: (a) they were working or studying in one (and only one) of the two fields, (b) they had high familiarity with the terms in their own field of expertise (top 40% compared to all participants), (c) they had low familiarity in the other field (bottom 50%), and (d) they did not consider themselves novices in the field (i.e. they did not rate their own familiarity with the field compared to other people working or studying in the field lower than 3 on a 7 point Likert scale). Participants' average familiarity with the concepts in the domain they were an expert in was higher than that of the other domain (see Table 7.1). In addition, each individual participant had higher familiarity with the terms from their own domain than with the terms from the other domain. Note that the biomedical experts have higher familiarity with the terms from the economics domain than vice versa. We will return to this issue in the discussion.

In short, the experts in our study had academic experience in the relevant subject (as shown by their registration on Prolific and their responses to our pretest) and considered themselves knowledgeable in the field (as indicated in our pretest) and their expertise was also reflected in their familiarity with the specialized language used in the texts.

Table 7.1: Mean scores on the concepts by domain and expertise. Scores on a scale from 1 (*I have not heard of the term*) to 3 (*I would be able to describe the term*).

	biomedical terms	economic terms
biomedical experts	2.86	2.09
economics experts	1.62	2.94

In the final experiment, 106 participants, all native speakers of English, took part (age range, 19-47 years; mean age, 24.7 years; 60 female). Of these, 89 participants were students; 56 had completed an undergraduate degree or had obtained a higher education level.

7.4.2 Materials

Ninety-six relations were sourced from the Penn Discourse Treebank 2.0 (PDTB Prasad et al., 2008), a discourse-annotated corpus containing Wall Street Journal texts. Only those sections from the PDTB that were classified as news articles were selected. In addition, we only included texts that covered economic or financial topics. An additional set of 96 relations was extracted from the Biomedical Discourse Relation Bank (BioDRB Prasad et al., 2011). This corpus contains discourse annotations of 24 biomedical research articles from the GENIA corpus, using an adapted version of the PDTB annotation framework. The latter texts are likely more specialized than the newspaper texts, which are written to be accessible to a broader audience. Still, financial newspapers, like biomedical research papers, target a specific group of readers (i.e. people working in the field of economics) and some degree of domain-specific knowledge is presumed by the writers of these texts as well. We will elaborate on this issue in the Discussion. Different items could come from (different parts of) the same texts, but the items in each corpus came from at least twenty different texts so that writing styles were varied.

The set of experimental items contained an equal amount of implicitated (i.e. originally explicit discourse relations from which the connective has been removed) and implicit relations. To balance the items with respect to the cognitive complexity and expectedness of the relation sense, four different relation senses were selected for the purposes of the present study: `result`, `contrast`, `concession` and `instantiation`. More specifically, we selected `contra-expectation` as the subcategory of `concession` relations.[4] Each relation sense occurred equally often in the experiment.

Only items for which both arguments were single full sentences were included. The context, consisting of one or two full sentences before and after the arguments, was also presented. An example of an implicitated `result` item can be found in Passage

[4]Within the class of contrast relations, the PDTB2 distinguishes between `juxtaposition` and `opposition`. Since no such distinction was made in the BioDRB, this distinction was disregarded when selecting materials for the present experiment.

57. The relational arguments, separated by \\ are presented in boldface here. To the participants, the context was presented in grey and the target sentences in black.

(57) Convertible debentures – bonds that can later be converted into equity shares – are the most popular instrument this year, though many companies are also selling non-convertible bonds or equity shares. These mega-issues are being propelled by two factors, economic and political. **In the past, the socialist policies of the government strictly limited the size of new steel mills, petrochemical plants, car factories and other industrial concerns to conserve resources and restrict the profits businessmen could make \\ industry operated out of small, expensive, highly inefficient industrial units.** When Mr. Gandhi came to power, he ushered in new rules for business. He said industry should build plants on the same scale as those outside India and benefit from economies of scale.

For each of the four relation senses and each relation marking, 12 items were extracted from the PDTB and 12 items from the BioDRB, resulting in 192 items in total.

7.4.3 Procedure

The task was an updated version of the two-step connective insertion task developed by Yung et al. (2019).[5] Participants were presented with each item one by one and were asked to complete two steps. In the first step, participants were asked to freely insert a connective in the blank that reflects the relation between the arguments best. They could only continue to the next step if they had typed something in the blank and were instructed to type the word *nothing* if they could not think of a linking phrase connecting the sentences.[6] They were then provided with a list of connectives in the second step and asked to select the connective that fits the relation best. The options presented in the second step were based on the insertion in the first step, and were unambiguous alternatives for the relations that can be signaled

[5]Three adaptations were made to Yung et al. (2019)'s task. Firstly, participants were always presented with the second step in this experiment, regardless whether the connective they inserted in the first step was unambiguous or not, to discourage the use of only very specific connectives in the first step. Secondly, the connective bank and the mapping of ambiguous connectives to the options in the second step was updated based on follow-up experiments. Thirdly, the default list was adapted for the purposes of the current study.

[6]Participants avoided this restriction in 1.2% of all data points by inserting punctuation or a whitespace.

by the connective in the first step. For example, if *but* was inserted in the first step, the options in the second step consisted of (among other options) *despite this* and *on the contrary* to disambiguate between the `concession` and `contrast` relation sense that can be marked by *but*. If the option inserted in the first step was not present in the connective bank, a default list was presented: *therefore, in addition, despite this, in more detail, even though, for example, by contrast, due to, this example illustrates that, in other words*. This default list thus contained a target connective for each of the target relations included in the item.

The experiment was hosted on Lingoturk (Pusse et al., 2016) and distributed via Prolific. Participants first received instructions. They then saw two practice items, after which they received feedback on possible answers for these items. The items were divided across three batches per relation marking, with four items per relation sense per domain. Thus, each participant saw 32 experimental items. Four additional filler items were included as attention checks. These items were taken from the PDTB and did not require economic domain knowledge. Performance for these items was at ceiling in previous experiments. After completing the study, participants were asked to rate the difficulty of the texts on economic and biomedical topics. The study took around 30 minutes to complete and participants were given £3.50 as compensation for their participation.

7.4.4 Data analysis

Data from participants who provided less than five different types of connectives in the first step (n = 2) or selected that they wanted to insert a different connective in the second step in more than half of the cases (n = 4) were excluded from further analysis. In addition, participants who failed to select a connective that belonged to the same relational class as the gold label (see below) for more than half of the filler items were also removed (n = 6). The final dataset (n = 2,976) contained observations of 48 experts in the domain of biomedical sciences (implicit: 23, implicitated: 25) and 44 experts in the domain of economics (implicit: 24, implicitated: 22). Trials for which participants answered that none of the options provided in the second step were suitable, were coded as missing data (n = 179 observations, 6.0%).

To determine whether participants had inferred the relation correctly, the connectives in the second step were categorized as signaling eight different relational classes: (1) cause, (2) temporal, (3) contrast, (4) concession, (5) positive expansion (e.g. `instantiation`), (6) negative expansion (e.g. `disjunction`), (7) condition, (8)

no relation. An overview of which PDTB3 relation senses are included in each relational class can be found in Appendix F.

We recoded a new variable, correctness, which was 1 when the inserted connective in the second step matched the relational class of the target relation sense (i.e. agreed with the gold standard), and 0 when it signaled a different relational sense. The correctness variable is used as the dependent variable in subsequent analyses, unless stated otherwise.

During data exploration, we discovered that performance on the `contrast` relations was much lower in the PDTB than in the BioDRB (11.2% vs. 31.9%) as well as compared to other relation senses in the PDTB (60.0%). For these contrastive PDTB items, participants frequently provided a concessive connective. Note that the distinction between `contrast` and `concession` is notoriously difficult (see e.g. Robaldo & Miltsakaki, 2014; Zufferey & Degand, 2017). In fact, the manual of the updated version of the PDTB2 (PDTB3 Webber et al., 2019) states that they addressed this issue in PDTB3 by reclassifying many `contrast` relations as `concession`. We compared the labels of our `contrast` items between PDTB2 and PDTB3 and found that 21 out of 24 `contrast` relations were relabelled as `concession`. We therefore decided to use the updated PDTB3 labels as the gold label.[7] We will come back to this issue in the discussion.

Binomial mixed-effects analyses were used to examine the data. Corpus, expertise and relation marking were deviation coded for ease of interpretation of the model with the PDTB corpus, economic experts and implicit relations at -1 and their counterparts at 1. For relation sense, treatment coding was used with `concession` as the intercept, as this was hypothesized to be one of the most difficult relations to infer. In addition, we were interested in its comparison with `contrast` relations, due to these relations often being confused. Because of convergence issues, the BOBYQA optimizer was used with 10,000 iterations. The models were always first constructed with maximal random effect structure. In case of non-convergence, the model was reduced (Barr et al., 2013). The random slope for relation sense never converged. Unless specified otherwise, the models therefore contained random intercepts for participants and items and random slopes for corpus and expertise.

[7]The label for `contra-expectation` and `instantiation` relations also differ between these two versions of the PDTB. These items were all labeled as `arg2-as-denier` and `arg2-as-instance` respectively in the PDTB 3.0.

Table 7.2: Confusion matrix of the gold relation senses and the inserted categories (% per relation sense).

	positive				
	cause	**expansion**	**concession**	**contrast**	**other**
`result`	64.7	17.5	8.7	2.7	6.3
`instantiation`	21.9	56.2	9.9	5.7	6.4
`concession`	19.1	15.0	48.8	8.1	9.0
`contrast`	15.0	25.7	21.2	31.4	6.7

7.5 Results

7.5.1 Convergence with the gold label

On average, the correct relation sense was inferred in 52.1% of the insertions, as shown by convergence of the insertions in the second step with the gold label. Although this performance is relatively low, discourse relation classification is a notoriously difficult task and these numbers are comparable with similar studies using crowd-sourcing for discourse relation annotation (cf. Rohde et al., 2016; Kishimoto et al., 2018; Scholman et al., 2022c). When the majority label per item is taken (i.e. aggregating responses of all participants per item to obtain a single annotated label), performance is much higher (74.2%).

As can be seen in Table 7.2, performance is much higher on some relation senses than on others (see Yung et al., 2019; Scholman & Demberg, 2017a, for similar results). Connective insertions for `result` items were correct in 64.7% of cases, followed by the `instantiation` relations (56.2%). These two relational classes were often confused, suggesting that participants did not always know whether the relation was causal or not. Another possibility is that these relations were ambiguous for these two relation senses, since `instantiation` relations can often also be causal (Scholman & Demberg, 2017b). `Concession` (48.8% correct) showed significantly lower accuracy than performance on `result` relations as shown in a binomial mixed-effects analysis (see Table 7.3 below). The difference with `instantiation` relations was not significant. `concession` relations were sometimes confused with `result`, but also with positive expansion relations. The latter is surprising, since that means that participants neither infer the causal nor the negative relation between the arguments. Finally, performance on `contrast` relations was even lower than on `concession` relations, with only 31.4% of insertions falling in the same category of the gold label.

In many of these items, a connective that signals positive expansion or `concession` was inserted. `Contrast` relations have been shown to be difficult to annotate in other studies as well (cf. Kishimoto et al., 2018). In addition, `contrast` and `concession` relations are known to often be confused with each other (e.g. Robaldo & Miltsakaki, 2014; Demberg et al., 2019). Given the effect of relation sense on accuracy, relation sense was included as a covariate in all models presented in this paper.

7.5.2 The effect of domain knowledge on discourse relation inferences

As can be seen in Figure 7.1, performance was higher in the PDTB (57.1%) than in the BioDRB (47.1%). This effect was confirmed in the model, as shown by a significant main effect of corpus (see Table 7.3).[8] In addition, overall, biologists converged with the gold label significantly more often than economists (54.8% vs. 49.4%). Indeed, expertise was also a significant predictor in the regression analysis.

The main question that this study aims to answer, however, is whether high-knowledge readers infer the correct relation sense more often than low-knowledge readers and how readers interpret discourse relations in the absence of domain knowledge. As can be seen in Figure 7.1, for each corpus, highest performance was obtained by the experts from that domain. The binomial mixed-effects analysis shows that the interaction between corpus and expertise is significant, suggesting that domain knowledge leads to a higher accuracy on relational inference. To examine this interaction more closely, we conducted pairwise comparisons. On items from the BioDRB, experts from the biomedical sciences converges with the gold label more often than economic experts (β = -0.62, SE = 0.18, z = -3.49, $p < .001$). For economic texts, this difference between experts and non-experts was not significant.

7.5.3 Interpretation strategies in the absence of domain knowledge

Exploiting discourse relational cues

We hypothesized that readers use discourse relational cues in the text to infer the relation. More specifically, we assumed that implicit relations contain more of these

[8] Model specification: `correctness ~ relationsense + corpus*expertise*relationmarking + (1 + corpus | workerid) + (1 + expertise | questionid)`

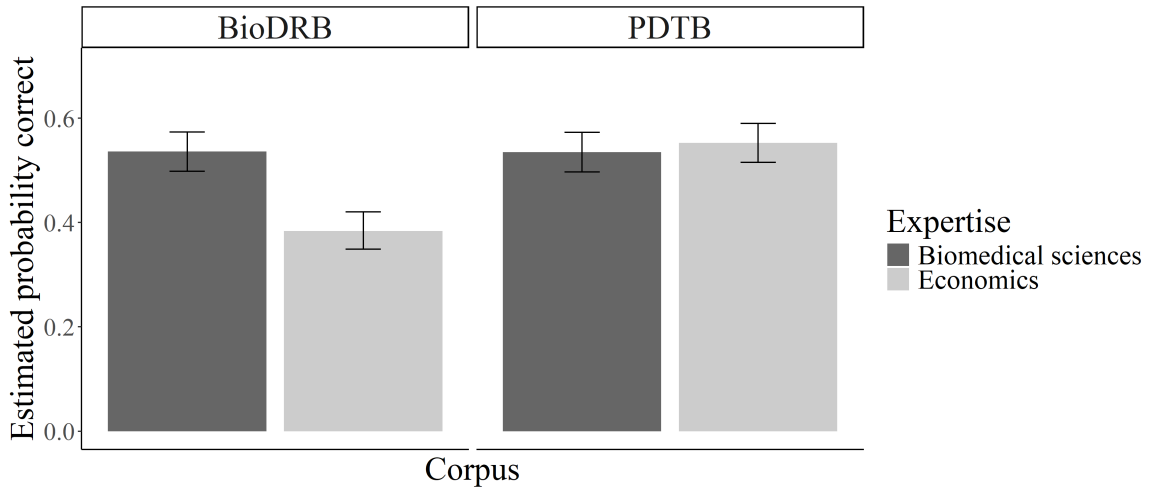

Figure 7.1: Estimated probability of convergence with gold label per domain and expertise with error bars showing the standard error.

Table 7.3: Model output on convergence with the gold label.

	β	SE	z	p
Intercept	-0.11	0.14	-0.81	.42
Relationsense `result`	0.84	0.20	4.20	**<.001**
Relationsense `contrast`	-0.75	0.26	-2.92	**<.01**
Relationsense `instantiation`	0.39	0.20	1.94	.05
Corpus	-0.17	0.08	-2.07	**.04**
Expertise	0.14	0.07	1.97	**.05**
Relationmarking	0.09	0.09	0.98	.33
Corpus:expertise	0.17	0.05	3.55	**<.001**
Corpus:relationmarking	0.04	0.08	0.54	.59
Expertise:relationmarking	-0.01	0.07	-0.18	.86
Corpus:expertise:rel...marking	-0.09	0.05	-1.92	.06

cues than implicitated relations and are therefore easier to infer. In addition, low-knowledge readers were hypothesized to rely on these cues more than high-knowledge readers and therefore perform better on implicit than on implicitated relations. However, relation marking was not a significant predictor for convergence with the gold label (see Table 7.3). There was also no three-way interaction between corpus, expertise and relation marking. We thus find no evidence that the implicit relations are easier to infer than implicitated relations, nor that the effect of domain knowledge is different in implicit and implicitated relations.

Possibly, however, the assumption that implicit relations contain more non-connective cues than implicitated relations is not correct. To examine the role of discourse relational cues more closely, we therefore performed a qualitative analysis on the 30 items for which the difference in accuracy between high-knowledge readers and low-knowledge readers was largest and examined the insertions by both groups. This allowed us to distinguish three different types of relations, which are presented below.

Relations without linguistic cues require domain knowledge For some items, the relation could only be inferred using domain knowledge. For instance, in Passage (58), a reader needs to know what 'Treg activities' are like in murine systems in order to know whether a reduction in human systems is similar or not. However, no linguistic cues are present to signal this relation. As a result, low-knowledge readers often interpreted this item as a `cause` relation, instead of `concession`.

(58) In human infectious, neoplastic, and autoimmune diseases, Treg activities often mirror those in murine systems ___ numbers of Treg are reportedly reduced in human autoimmune diseases, (...)

(BioDRB:`concession`:Implicit)

Relational cues allow relational inferences in the absence of domain knowledge A number of items on which experts and non-experts diverged, contained non-connective cues, such as hyponyms for `instantiation` relations and antonyms for `contrast` relations. These could help readers to infer the correct relation (cf. Crible & Demberg, 2020). More specifically, the majority of the fourteen `instantiation` and `contrast` items that yielded a large difference between experts and non-experts contained such a cue. To illustrate these cues, consider (59) and (60), which yielded high accuracy from both high- and low-knowledge readers. The relational cues in these items are signaled linguistically by repeating words (e.g. *magazine*) or are based on

general world knowledge contrasts (left vs. right). This allows readers to infer relations even in the absence of domain knowledge.

(59) Other **magazine publishing companies** have been moving in the same direction ___ **the New York Times Co.'s Magazine Group** earlier this year began offering advertisers extensive merchandising services built around buying ad pages in its Golf Digest magazine.

(PDTB:`instantiation`:Implicit)

(60) The core biopsy of the **left breast** revealed infiltrating ductal carcinoma in 2 of 5 core fragments; high nuclear grade, with no lymphatic invasion seen ___ the core biopsy of the **right breast** demonstrated benign pathology, specifically, fibrosis with focal ductal epithelial hyperplasia.

(BioDRB:`contrast`:Implicit)

Relational cues sometimes require domain knowledge In the items where there was a large difference between experts and non-experts, low-knowledge readers did not always pick up on these cues. The reason for this is that domain knowledge was often required to exploit the cue. This was especially the case for the `instantiation` relations. In about half of the cases in which a hyponym was present, this cue could only be exploited with domain knowledge. For example, in (61) below, the reader needs to know that *orthologous genes* are genes in different species that have a similar descent. The second argument provides a specific example of this, but if a reader does not have the required domain knowledge, they will likely also not understand that these genes are instances of orthologous genes.

(61) In particular, we assumed that the transcriptional regulation is conserved for **orthologous genes ___ the mouse gene Myh1 and the human gene MYH1** are assumed to share expression patterns and to share important cis-regulatory sequences.

(BioDRB:`instantiation`:Implicit)

Interestingly, the largest difference between experts and non-experts in convergence with the gold label in the full dataset can be found in implicitated `instantiation` relations in the BioDRB. Experts performed 30 percentage points higher than non-experts in this condition (see Table 6 in Appendix G). The implicit `instantiation` items in the BioDRB and implicitated `instantiation` items in the PDTB also yielded higher accuracy for experts than for novices. This suggests that cues for

`instantiation` relations are more easily exploited by experts. In a post-hoc analysis, we therefore examined whether the effect of domain knowledge was different per relation sense. Adding the three-way interaction between relation sense, corpus and expertise did not significantly improve model fit when compared to the same model without this interaction. Since examining differences between the relation senses was not the purpose of the present study and power for finding such a three-way interaction effect with the current study design is likely to be low, further quantitative research is necessary to examine the effect of domain knowledge on different relation senses and different relational cues. The present qualitative analysis provides directions for future research.

Furthermore, it is interesting to point out that low-knowledge readers do not always exploit relational cues that do not require domain knowledge. More specifically, the three antonyms in the `contrast` relations that were more challenging for low-knowledge readers could also be detected with general world knowledge, contrasting concepts that are accessible for low-knowledge readers as well (see (63) for an example). In addition, we found instances of hyponyms in our qualitative analysis that do not require specific domain knowledge to infer the `instantiation` relations, but were nevertheless not detected by low-knowledge readers, as in (62).

(62) More recently, **several groups** have demonstrated the feasibility of hybridizing metabolically labeled mRNAs directly from nuclear run-on (NRO) reactions to nylon filter microarrays in order to investigate nascent transcripts ___ **Schuhmacher et al.** used a B cell line carrying a conditional, tetracycline-regulated myc gene, and found that myc induction resulted in only a small overlap in regulated mRNAs at 4 hours post-induction when comparing polyA mRNA and NRO RNA on microarrays.

(BioDRB:`instantiation`:Implicitated)

(63) E2 **inhibits** apoptosis in different cell types (cardiac myocytes and others) ___ androgens have been found to **induce** apoptosis.

(BioDRB:`contrast`:Implicitated)

To sum up, non-connective cues seem to play a role in discourse relation inferences, although we do not find evidence that the presence of these cues (or the extent to which they are used to infer the discourse relation) depends on whether or not the relation is marked. In addition, the qualitative analysis shows that adopting these cues sometimes requires domain knowledge. However, even if domain knowledge is not required, low-knowledge readers do not always adopt these cues.

Table 7.4: Model output on insertion of causal and continuous connectives.

	β	SE	z	p
Intercept	1.68	0.21	7.89	**<.001**
Relationsense `result`	-0.69	0.31	-2.22	**.03**
Relationsense `contrast`	-0.96	0.36	-2.67	**<.01**
Relationsense `instantiation`	-1.05	0.30	-3.45	**<.001**
Corpus	0.28	0.12	2.23	**.03**
Expertise	0.12	0.09	1.33	.18
Relationmarking	0.14	0.12	1.13	.26
Corpus:expertise	0.03	0.08	0.37	.71
Corpus:relationmarking	0.29	0.12	2.48	**.01**
Expertise:relationmarking	-0.04	0.09	-0.45	.65
Corpus:expertise:rel...marking	0.14	0.08	1.82	.07

Cognitive bias for continuity and causality

A second hypothesis was that readers might be guided by cognitive biases for causality and continuity in case their background knowledge was insufficient to determine the relation sense. To examine whether low-knowledge readers resorted to default interpretation strategies, we coded the connective insertions for whether they were signals of continuous relations (*cause, positive expansion, temporal, condition*) or not (*contrast, concession, negative expansion)*. We only included incorrect insertions in this analysis, because correct continuous or discontinuous interpretations are likely guided by the true sense of the relation. A binomial mixed-effects regression analysis[9] revealed that default interpretations were more likely for `concession` relations than in the other relations, as shown in Table 7.4. Furthermore, there was a significant interaction between corpus and relation marking, showing that continuous interpretations were less frequent in implicit relations in the BioDRB (β = -.085, SE = 0.34, z = -2.53, p = .01). Crucially, however, there was no interaction between corpus and expertise. In other words, we find no evidence that readers resort to default interpretation strategies in the absence of domain knowledge.

[9]Model specification: `defaultint` $\sim$ `relationsense + corpus*expertise*relationmarking + (1 | workerid) + (1+expertise|| questionid)`

Table 7.5: Model output on the insertion of ambiguous connectives in the first step.

	β	SE	z	p
Intercept	0.34	0.11	2.96	**<.01**
Relationsense `result`	-0.99	0.15	-6.57	**<.001**
Relationsense `contrast`	0.06	0.19	0.33	.74
Relationsense `instantiation`	-1.48	0.16	-9.42	**<.001**
Corpus	-0.08	0.06	-1.35	.18
Expertise	-0.15	0.08	-1.94	.05
Relationmarking	-0.07	0.08	-0.87	.39
Corpus:expertise	-0.14	0.04	-3.24	**<.01**
Corpus:relationmarking	-0.09	0.06	-1.63	.10
Expertise:relationmarking	-0.04	0.08	-0.54	.59
Corpus:expertise:rel...marking	0.01	0.04	0.20	.84

Underspecified interpretations

If readers are not certain about the discourse relation between two arguments, they might resort to making an underspecified inference, rather than committing to a specific interpretation. In the paradigm used in this experiment, this would mean that participants insert more ambiguous connectives in the first step when they have little knowledge about the domain of the text, compared to when they are experts in that domain. The connective insertions in the first step were therefore annotated as indicating relations from one vs. multiple relational classes. The most frequent ambiguous first step insertions were *however* (11.8%), *and* (6.5%) and *but* (5.9%). In addition, participants typed *nothing* in 3.4% of cases, which indicated that they could not come up with a linking phrase connecting the sentences. *For example* (6.4%), *therefore* (5.1%) and *because* (3.3%) were the most frequent specific connectives.

A binomial mixed-effects logistic regression analysis showed an interaction between corpus and expertise (see Table 7.5). This effect of domain knowledge on the insertion of ambiguous connectives is visualized in Figure 7.2. Pairwise comparisons revealed that economic experts inserted significantly fewer relation-specific connectives in the BioDRB (β = -0.57, SE = 0.17, z = 3.26, $p < .01$). No effect was found for the PDTB. Besides inferring more incorrect relation types, low-knowledge readers thus also leave the relation underspecified by inserting ambiguous connectives in the first step, when reading the BioDRB.

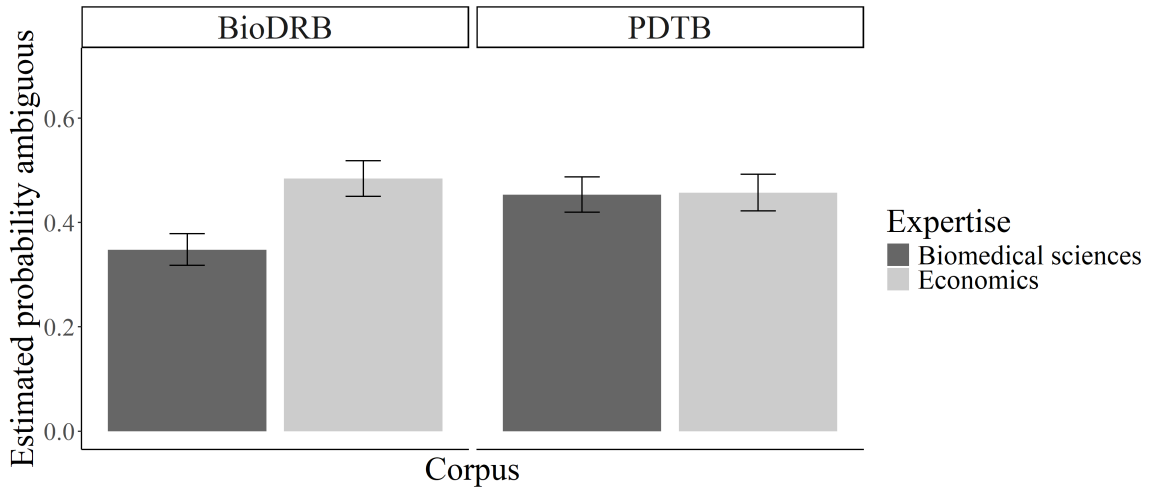

Figure 7.2: Estimated probability of ambiguous insertions in the first step with error bars showing standard error.

7.6 Discussion

Background knowledge has often been assumed to play a role in correctly interpreting discourse relations, but this has never been investigated experimentally. The current study filled this gap by assessing discourse relation interpretations of high- and low-knowledge readers. We aimed to examine whether domain knowledge contributes to inferring the correct discourse relation, as well as which factors guide discourse relation interpretation in the absence of connectives and domain knowledge. The first main finding of this research is that high-knowledge readers were better at inferring the discourse relation, as measured by convergence with the gold label, than low-knowledge readers. Thus, domain knowledge can, in some instances, facilitate establishing coherence and readers are able to employ their knowledge base to interpret the relation correctly. However, this effect was modulated by the corpus from which the text was taken: The effect of expertise was significant for the items from the BioDRB, but not for the PDTB (see Section 7.6.1). In addition, we identified non-connective linguistic signals for discourse relations, showing that domain knowledge influences how readers adopt these cues (see Section 7.6.2).

7.6.1 Text genre and the influence of domain knowledge

One possible reason for why there was only an effect of domain knowledge for texts from the BioDRB and not the PDTB is the difference between these specific genres. Even though economics newspaper texts are targeted at readers with a specific in-

terest in economics, they are intended for a broader audience with various levels of expertise. Research papers, on the other hand, are often not accessible to a general audience. Instead, they specifically target experts in that domain. They contain more specialized vocabulary and focus on topics that only a limited amount of people are familiar with.[10] Also note that the two texts differ in that they are written by journalists, who are not experts themselves, versus researchers. Discourse relations in the biomedical texts therefore likely required more domain knowledge than those in the economics newspaper texts.

Another explanation for this pattern could be related to the level of expertise of our participants. We recruited the participants via a crowd-sourcing platform, but their expertise was assessed in various ways (among others their subject of study as indicated on Prolific and their familiarity with specialized terms as determined during prescreening). This ensured that they were indeed high vs. low-knowledge readers with respect to the texts presented in this study. We note here again that experts were not expected to be familiar with all the information in the text. Domain knowledge was hypothesized to help in interpreting discourse relations correctly, because text processing is facilitated by an existing knowledge structure. This knowledge base does not need to be exhaustive, as the information in the text fills gaps in existing knowledge. Still, the experts from the domain of biomedical sciences seemed to know more about economics than vice versa, as measured by their self-rated familiarity with specialized terms from texts from that domain. They therefore might have also been able to rely on their background knowledge of some economic topics, when interpreting the relations. The finding that the effect of domain knowledge is smaller for the items from the PDTB can therefore not be considered surprising.

This interaction also raises the question of what constitutes expertise. Experts were assumed to be more knowledgeable with respect to the topic of the text. This knowledge would have been gained through reading texts typical to the domain. In the case of the biomedical experts, this would more likely be research papers than newspapers; in the case of economics experts, this would more likely be economic newspapers than research papers. In our post-test questionnaire, biomedical experts indeed indicated that they read research papers more often than economists (mean 3.57 vs. 2.59 on a 1-5 Likert scale ranging from 'never' to 'daily'). This difference was even more distinct for biomedical research papers (3.41 vs. 1.30). Economics experts,

[10]This was also confirmed in a post-hoc analysis of the perplexity of the items using a generic language model (GPT by Radford et al., 2018, a transformer model that is trained on a 1B word book corpus). The perplexity of the PDTB items (89.0) is lower than of the BioDRB items (101.4).

on the other hand, read newspapers (and specifically business newspapers) more often than biomedical experts (3.5 vs. 2.35 for newspapers in general; 3.09 vs. 1.33 for business newspapers). Domain knowledge might thus consist not only of topic knowledge, but also of text genre familiarity. Such familiarity might help readers to infer the discourse relations in that genre. For example, newspaper texts are characterized by a so-called *inverted pyramid scheme*, where the first paragraph is followed by an `elaboration` in the subsequent paragraphs (Das et al., 2018). Methodology sections of research papers often contain many `temporal` relations (Bachand et al., 2014). Even readers who are not familiar with the domain of the text (e.g. psycholinguistic researchers when reading biomedical research papers) might use genre familiarity with the text structure to infer the discourse relation. A future line of research could attempt to further tease apart the influence of topic knowledge and genre familiarity on the impact of domain knowledge, and how these two factors separately contribute to inferring discourse relations.

7.6.2 Discourse relational cues

Besides examining the role of domain knowledge in inferring the correct discourse relation, we also set out to explore how readers infer discourse relations in the absence of domain knowledge. The first prediction was that non-connective linguistic cues for discourse relations would be used. We therefore varied whether a connective was present in the original text. Discourse relational cues were assumed to be more frequent in implicit than in (originally explicit) implicitated relations (cf. Sporleder & Lascarides, 2008; Hoek et al., 2019b; Crible, 2020), which is why we expected implicit relations to be easier to infer than implicitated relations. However, no effect of the presence of a connective in the original text was found. In addition, high- and low-knowledge readers were not affected differently by whether the relation was originally marked. We can therefore not confirm the hypothesis that discourse relational cues in implicit relations facilitate discourse relational inferences.

The by-item analysis revealed that some discourse relations contained cues for the relation, even in the absence of a connective. For example, antonyms were present in `contrast` relations and hypernyms in `instantiation` relations. Low-knowledge readers sometimes successfully retrieved the relation when such cues were present. However, they were not always sensitive to these cues. For example, hypernyms did not always help low-knowledge readers to infer an `instantiation` relation, nor did antonyms in `contrast` relations. Instead, readers strongly diverged in the interpretations of these items.

There are several possible explanations for these findings. First of all, non-connective linguistic signals for discourse relations are highly ambiguous, with many signaling a large variety of discourse relations (see also Chapter 6). The cue might then exclude some possible relation interpretations, but not provide only one single likely interpretation. As a result, different readers might interpret the cue differently, diminishing the facilitative effect of additional relational cues in implicit relations. Secondly, signals for discourse relations (other than connectives) may require domain knowledge to interpret them. To illustrate, antonyms could serve as a signal for `contrast` relations, but to know that two concepts are opposite, the reader should know what the words mean. This could explain why low-knowledge readers do not always pick up on these cues. However, even when only general world-knowledge was required to interpret the cue, low-knowledge readers still did not always exploit these signals. Possibly, readers use non-connective cues to confirm their interpretation, rather than to explore different relation senses. To illustrate, a high-knowledge reader might consider a particular relation (e.g. `contrast`) based on their domain knowledge and subsequently exploit antononomy as additional evidence for this relation. A low-knowledge reader, however, cannot use their domain knowledge to arrive at an initial interpretation and subsequently also does not recognize linguistic cues to confirm such an interpretation. This confirmatory role of non-connective cues also suggests that readers rely more heavily on the content of the relation than on the cue. This would explain why the effect of non-connective signals seems limited in tasks in which the discourse relation can be inferred based on the content (cf. Grisot & Blochowiak, 2017; Crible & Pickering, 2020, see also Chapter 6, Experiments 2 and 3), except when readers have to disambiguate between one of four relations (Crible & Pickering, 2020) Furthermore, in tasks in which the content of the discourse relation is not provided, findings on the role of non-connective cues are more consistent: non-connective cues have repeatedly been shown to elicit expectations for upcoming discourse relations (cf. Kehler et al., 2008; Rohde & Horton, 2014; Scholman et al., 2020; Bott & Solstad, 2021, see also Chapter 6, Experiment 1). However, further research is needed to confirm this hypothesis that the non-connective cues are used to confirm discourse relation inferences.

Another explanation for the diverging interpretations of items containing non-connective linguistic signals might lie outside the scope of the text itself and be influenced by characteristics of the reader. Scholman et al. (2020) show that some readers are more sensitive to contextual list signals than others. More specifically, participants in their study who had more reading experience (as measured by an Author Recognition Test), picked up on these cues more than participants who were less

experienced readers. Only some readers might therefore have been able to employ these signals in inferring the relation, leading to differences in how relations containing such cues are interpreted. With respect to domain knowledge, high-knowledge readers had access to two strategies in interpreting the relation: non-connective linguistic signals and their knowledge base. The high-knowledge readers who were not sensitive to these non-connective discourse relational cues could then use their domain knowledge to infer the relation, whereas low-knowledge readers would not be able to interpret the relation if they did not detect these signals.

7.6.3 Inferences in the absence of domain knowledge

The present study also set out to investigate what low-knowledge readers do when they lack the domain knowledge that is required to infer the correct discourse relation. Apart from using non-connective linguistic signals, we hypothesized that participants might resort to a default interpretation strategy and have a preference for causal and continuous relation interpretations in cases in which the relation was not inferred correctly. However, low-knowledge readers did not insert causal and continuous connectives in incorrect items to a greater extent than high-knowledge readers. We thus did not find evidence that domain knowledge influences readers' cognitive biases for causality and continuity.

In addition, we predicted that low-knowledge readers would prefer to leave the discourse relation underspecified. Rather than committing to a certain interpretations that might be incorrect, readers were hypothesized to make underspecified discourse interpretations and therefore provide connectives reflecting this underspecification. We found some evidence for this hypothesis, since low-knowledge readers inserted more ambiguous connectives in the first step than high-knowledge readers. Low-knowledge readers thus seem to avoid making a specific relation interpretation. However, it remains unclear what the reason for these underspecified interpretations is. On the one hand, it is possible that low-knowledge readers were unable to specify the relation further. On the other hand, low-knowledge readers might have processed the text less deeply and therefore not committed to a specific relation because they did not wish to do so. Future research could examine whether low-knowledge readers perform better when they are forced to process the text more deeply (cf. Scholman, 2019) to disentangle these two factors.

7.6.4 Limitations

Finally, we note some limitations of the present research. Firstly, the study aimed to balance the items among the different discourse relation senses, since different relation senses were hypothesized to yield differences in accuracy and interpretation biases. We did indeed find that `result` and `instantiation` relations were easier to infer for participants than `concession` relations and that `instantiation` and `concession` relations were often interpreted as being causal. Many of the initially selected `contrast` relations had been annotated as `concession` in the PDTB 3. The lower performance on this relation sense could therefore partly be attributed to the disagreement about the gold label, since a `concession` interpretation might also have been possible. However, since relation sense was included as a covariate in the analysis, this does not affect the conclusions about the role of domain knowledge.

Another limitation is our manipulation of relation marking. It is possible that the relation might have become impossible to identify or has changed by removing the connective. In the first case, we would find floor effects on the implicitated relations, even for the high-knowledge readers. Overall, there were ten (out of 190) items for which none of the high-knowledge readers converged with the gold label. However, these were equally distributed over the implicit and implicitated condition. This suggests that the original relation could still be retrieved, even when the connective had been removed, also in the implicitated condition. In the case of multiple interpretations, convergence to the gold label is not reliable anymore. To account for the problem of multiple interpretations, we examined those implicitated relation items where several participants agreed on the same non-gold relation interpretation and assessed whether this interpretation was also possible. Including these alternative answers as correct still revealed the same pattern as above: high-knowledge readers interpreted the relation correctly more often than low-knowledge readers. Nevertheless, a study manipulating discourse relational cues specifically would provide further insight on this matter.

Furthermore, despite carefully selecting our participants, we cannot be sure that they were indeed as knowledgeable as they said they were (but see Mehti, 2024, for evidence that self-assessments correlates with performance on textbook questions). Nevertheless, we found a clear effect of domain knowledge in the BioDRB, suggesting that the biomedical experts were indeed more familiar in this domain than the economic experts. There is no reason to believe that the experts from the field of economics would be less knowledgeable than the participants from the biomedical domain.

7.7 Conclusion

The goal of this chapter was to examine how reader characteristics influence the interpretation of discourse relations and their sensitivity to linguistic signals (cf. Research Goal 3, Chapter 1). Previous work has mainly focused on the influence of domain knowledge on text comprehension and recall (e.g. McNamara et al., 1996; Smith et al., 2021) or on whether or not discourse inferences are made in the absence of domain knowledge (e.g. Noordman & Vonk, 1998), showing that low-knowledge readers benefit more from coherence marking than high-knowledge readers and are less likely to make relational inferences during reading. However, these studies did not address how the discourse is interpreted differently by high- and low-knowledge readers. The present study shows that readers are able to interpret discourse relations correctly, even if they have little knowledge about the domain of the text. Still, high-knowledge readers make more correct (and more specific) discourse relation interpretations. This effect was established in biomedical research papers, a text type that targets a specialist audience, but not in economic newspapers, possibly because the genre is aimed to be accessible for both experts and novices in the field. Moreover, we found that readers adopt linguistic cues for inferring discourse relations, although this did not interact with the presence of a connective in the original text. A text without discourse connectives is therefore not necessarily detrimental for low-knowledge readers (cf. McNamara et al., 1996), as they can also establish coherence with other discourse cues. Still, these cues might be more challenging to low-knowledge readers as in some cases domain knowledge is required to detect them. Finally, our findings suggest that non-connective cues might be used to confirm rather than to explore discourse relation interpretations, but further research is needed to confirm this hypothesis.

Data availability

All materials, data and scripts can be found online: `https://osf.io/gq59w/`

Funding

This work was supported by the German Research Foundation (DFG) under Grant 232722074 - SFB 1102 ("Information Density and Linguistic Encoding"). The second and third author were supported by the European Research Council under Grant 948878 ("Individualized Interaction in Discourse").

Dissemination

The following journal article is based on this research:

- **Marchal, M.**, Scholman, M.C.J., & Demberg, V. (2022). The effect of domain knowledge on discourse relation inferences: Relation marking and interpretation strategies. *Dialogue & Discourse*, 13(2), 49-78.

In addition, (parts of) this research have been presented at the following conferences:

- **Marchal, M.**, Scholman, M.C.J., Demberg, V. (2023). The effect of domain knowledge on discourse relation inferences: Relation marking and interpretation strategies. *Poster at SigDial, Prague, Czech Republic,* 11-15 September.

Part IV

General discussion and conclusion

Chapter 8

General discussion

Discourse comprehension is a key component of successful communication. In order to understand a text, whether written or spoken, comprehenders need to make sense of how the different parts of the discourse are related. These discourse relations can be derived based on various sources of information. Previous research has mainly focused on the role of connectives in building a coherent mental representation of a text. Much less is known about how other linguistic signals influence the processing and representation of discourse relations. This dissertation therefore set out to investigate the following research question:

Which factors influence readers' sensitivity to linguistic signals of discourse relations?

In this dissertation, we investigated four factors influencing reader's sensitivity to signals of discourse relations: characteristics of the signal, the relation, the reader and the language. Here, we summarize the findings from the studies presented in this dissertation, as well as from earlier work, to provide an answer regarding each characteristic.

8.1 Characteristics of the discourse signal

Previous research has often focused on the role of connectives in discourse processing and representation (e.g. Cozijn et al., 2011; Kleijn et al., 2019; Asr & Demberg, 2020), but much less is known about the effect of other discourse signals. In Chapter 3, we

show that discourse relation signals differ with respect to five features. *Functionality* refers to whether the signal's primary function is to convey how segments are related. To what extent the signal is informative in determining the type of discourse relation is captured in the *informativity* feature. *Immutability* relates to whether the signal can be modified and *lexicality* to whether it is lexical. *Agreement* concerns the similarity between the signal's primary meaning and that of the discourse relation. These features are also argued to affect their effectiveness in signaling the discourse relation. The functionality, lexicality and immutability features describe how easily a signal can be identified: Immutable, lexical items that do not have another function can be recognized more easily as a discourse signal. The *informativity* and *agreement* feature determine how informative the signal is about the relation: if the cue always signals the same relation or if its propositional meaning overlaps with that of the relation, the relation can be derived more easily. Thus, we hypothesized that readers will be more sensitive to signals that adhere to these features more strongly. In the next section, we discuss the findings from our case study on the role of gerund free adjuncts in the processing of `result` relations. This is followed by a reflection on how readers' sensitivity to discourse signals is influenced by the characteristics described above (Section 8.1.2) as well as other factors suggested in previous literature (Section 8.1.3).

8.1.1 Gerund free adjuncts as a discourse relation signal

To extend research on signals other than connectives, we examined a cue that is not specialized for signaling discourse relations in Chapter 6. We showed that gerund free adjuncts (GFAs) provide information about the discourse relation through their frequent co-occurrence with `result` relations in Section 6.4. However, we also found that it can occur with almost any discourse relation, suggesting its informativity is low. In addition, as illustrated in Table 8.1, this cue is non-lexical and its overt form is context-dependent. Furthermore, its primary meaning is not similar to that of the discourse relation. Based on the hypothesis that weaker adherence to these features lead to a weaker influence on processing and representation, one might expect that the role of GFAs in discourse relation inference is limited.

We found evidence that readers adjust their off-line expectations about upcoming discourse relations based on these cues: participants provided more `result` relations following a GFA compared to a full matrix clause in Experiment 1 (Section 6.5). However, these expectations about the discourse relation did not influence the processing of `result` relations. More specifically, `result` relations were not read faster in the

Table 8.1: Overview of example discourse signals with their features, as well as their effect on representation or processing. For the features, black bullets indicates that the feature is (strongly) present. Gray bullets show that the feature is less strong or absent. For the findings, black squares indicates that there is strong evidence for the effect of this type of discourse signal, with ■ indicating smaller effects or mixed findings and ■ representing that no effect has been found. -: not applicable

	because; nevertheless	*and; but; when*	*this caused; due to rain*	*several; true*	*not*; antonyms	implicit causality	tense; eventuality	parallelism	gerund free adjuncts
functionality	●	●	●	●	●	●	●	●	●
informativity	●	●	●	●	●	●	●	●	●
immutability	●	●	●	●	●	●	●	●	●
lexicality	●	●	●	●	●	●	●	●	●
agreement	-	-	-	●	●	●	●	●	●
representation	■	■	■	■	■	■	■	■	■
processing	■	■	■	■	■	■	■	■	■

presence of a GFA compared to a full matrix clause. This is contrary to connectives (e.g. Cozijn et al., 2011; Blochowiak et al., 2022, see also Chapters 4 and 6), which facilitate the processing of relations they signal. This suggests that the role of GFAs in processing discourse relations is limited. Furthermore, we also did not find that readers have a preference for GFAs compared to full matrix clauses in `result` relations. These findings thus support the hypothesis that readers are less sensitive to signals that are less informative, non-lexical, modifiable, and whose primary meaning is not in agreement with that of the discourse relation.

8.1.2 A continuum of discourse signals in processing and representation

Which of these features influences reader's sensitivity to discourse relations? In Chapter 3, we show that functionality, informativity and lexicality seem to influence the

extent to which a signal affects processing and representation. Signals whose primary function is to signal the discourse relations (e.g. connectives) elicit stronger discourse expectations and facilitate processing more than signals that have a different primary function (e.g. quantifiers, tense, cf. Scholman et al., 2017; Cozijn et al., 2011; Scholman et al., 2020; Grisot & Blochowiak, 2017). Less informative (i.e. ambiguous) connectives have a weaker effect on processing than relation-specific connectives (Cain & Nash, 2011; Li et al., 2017). Similarly, lexical cues for discourse relations (e.g. implicit causality verbs, antonyms) have been shown to facilitate discourse processing and representation, but studies on non-lexical cues (e.g. tense, parallelism) often find limited effects (cf. Hoek et al., 2021a; Crible & Demberg, 2020; Grisot & Blochowiak, 2017; Crible & Pickering, 2020). For *immutability* and *agreement*, the findings on whether they affect readers' sensitivity are mixed. Both immutable connectives and context-dependent signals facilitate the representation and processing of relations they signal (cf. Van Silfhout et al., 2015; Sanders & Noordman, 2000). In addition, although signals that are in agreement with their relational meaning (e.g. negation, antonyms) seem to influence processing more than those that are not (e.g. parallelism, gerund free adjuncts, cf. Crible, 2021; Crible & Demberg, 2020; Crible & Pickering, 2020, Chapter 6), this is not the case for implicit causality verbs. There is evidence that these verbs have a strong influence on the representation and processing of discourse relations (cf. Kehler et al., 2008; Rohde & Horton, 2014; Hoek et al., 2021a), despite their primary meaning not being in agreement with the discourse relation. One explanation could be that there are other discourse signal features, such as at-issueness, that moderate readers' sensitivity to these signals. We will return to this in the next chapter (Section 9.1.1).

Note, however, that discourse signals' adherence to the features often overlaps. To illustrate, specialized discourse cues are often also immutable and more informative in comparison to non-specialized signals. It is thus unclear if all of these characteristics independently affect the extent to which readers use the signal when inferring the relation. To examine whether a feature independently influences readers' sensitivity, all other factors need to be kept constant. For example, ambiguous connectives differ from relation-specific connectives only with respect to their informativity. This shows that informativity independently influences readers' sensitivity. For the other features, more research is needed to compare the role of the other features in the processing and representation of discourse relations. We will discuss directions for further research in more detail in Chapter 9.

8.1.3 Other signal characteristics

The characteristics of the discourse signal presented here are not exhaustive. Other features of discourse signals have also been shown to influence whether readers use them to infer discourse relations. We will discuss two of these features here. Note that they were not included in the classification presented above, as they do not clearly distinguish between connectives and non-connective cues.

Scholman et al. (2024a) find that lexical transparency, the extent to which the meaning of a word can be derived from its form, influences performance on a coherence judgment task: Connectives that are more lexically transparent were found to be easier to comprehend. Lexical transparency thus likely also affects the extent to which readers will rely on a discourse signal in discourse interpretation. This notion is similar to that of *agreement* for non-connective cues: it quantifies the extent to which the discourse relation meaning can be derived from the signal itself.

Frequency is another factor that has been argued to be important in the comprehension of connectives. A number of studies show that readers comprehend more frequent connectives better than infrequent ones (Nippold et al., 1992; Zufferey & Gygax, 2020a,b; Tskhovrebova et al., 2022a, but see Wetzel et al., 2020; Scholman et al., 2024a). This suggests that readers are more sensitive to more frequent connectives. Indeed, readers detect the inappropriate use of connectives earlier when they are more frequent (Wetzel et al., 2021). With respect to non-connective cues, frequency might also influence readers' sensitivity. In a statistical learning account of discourse signals (cf. Chapter 6), not only the co-occurrence of the signal with the meaning should play a role, but also its general occurrence frequency. Even if a discourse signal is highly informative (e.g. the signal always occurs with the same relation), a language user needs repeated exposure to the signal to acquire the signal∼meaning mapping.

8.2 Characteristics of the discourse relation

In addition to characteristics of the discourse signal, we investigated how characteristics of the discourse relation influence the effect of discourse signals on on-line processing. Some discourse relations are more difficult to process than others (e.g. Sanders & Noordman, 2000; Noordman & de Blijzer, 2000; Xu et al., 2018). In addition, the representation of some discourse relations is stronger than of others, resulting in better memory of these relations (Myers et al., 1987; Sanders & Noordman, 2000; Trabasso & Van Den Broek, 1985). In Chapter 7, we showed further evidence for

relation-dependent differences in discourse processing: in the absence of a connective, *result* relations are inferred more accurately than `contrast` and `concession` relations. We hypothesized that features of the discourse relation would affect to what extent readers rely on connectives in discourse processing. More specifically, we predicted that readers would benefit from connectives more in discourse relations that are more difficult to process. We investigated two factors that influence the processing difficulty of the relation: whether the relation is causal or not, discussed in Section 8.2.1 below, and whether it is predictable or not, discussed in Section 8.2.2.

8.2.1 Causal vs. non-causal relations

To investigate whether the causality of the discourse relation influences readers' sensitivity to the presence of a connective, we compared the effect of the presence of a connective on the processing of `result` vs. `contrast` and `consession` relations in Chapter 4. `Result` relations were hypothesized to be easier to process, as readers have been argued to assume a causal relation by default (*causality-by-default* hypothesis). We therefore expected that the presence of a `result` connective should not provide as much benefit as a connective of a relation that readers do not infer automatically (i.e. `contrast` or `concession`). In Experiment 1, we compared the effect of connective presence in `result` relations to `contrast` relations. Surprisingly, we did not find any effect of relation marking on neither the region where the relation became clear nor in the subsequent clause-final region. In Experiment 2, the presence of a connective did facilitate the processing of the sentence-final region, but only for `concession` relations. This confirmed our hypothesis that readers are more sensitive to the presence of a connective in non-causal relations.

Note that our findings do not provide conclusive evidence for the causality-by-default hypothesis. First of all, we do not find a general advantage for `result` relations compared to `concession` relations. The effect of relation in Experiment 1 of Chapter 4 could also be attributed to differences in clause structure. Furthermore, a strong version of the causality-by-default hypothesis might predict that there is no facilitating effect of causal connectives at all. However, in Chapter 5, we show in both a self-paced reading study as well as an eye-tracking study that the presence of a `result` connective leads to shorter reading times of the material following it.

The effect of connectives in causal and non-causal relations also provides insights into the time-course of discourse relation processing. In sentence-initial regions, processing of both causal and non-causal relations has been found to be facilitated in the presence of a connective (Millis & Just, 1994; Cozijn et al., 2011; Van Silfhout

et al., 2014, 2015, see also Chapter 5). This effect has been attributed to facilitated propositional integration in the presence of a connective: the connective helps readers to understand how the two segments are related. This process of propositional integration extends to sentence-final regions as well in non-causal relations. The findings in Experiment 2 of Chapter 4 show that the presence of a non-causal connective facilitates sentence-final reading as well (see also Zufferey & Gygax, 2016). For causal relations, the presence of a causal connective does not facilitate reading here. In fact, some studies even find a negative effect of relation marking on sentence-final processing (Millis & Just, 1994; Cozijn et al., 2011). This suggests that readers might have already finished the integration process before the end of the sentence, possibly due to assuming a causal relation by default.

8.2.2 Predictability of the discourse relation

In Chapter 5, we investigated whether the processing difficulty of a discourse relation can be quantified in terms of its predictability. More specifically, we examined two levels of predictability: that of the relation type, operationalized as *relation surprisal*, and that of the content, operationalized as *semantic information value*. The second argument is processed faster when its content is more predictable. In addition, we show that the more predictable the relation type is, the better readers can predict the content. As a result, the predictability of the relation type facilitated reading by enhancing the predictability of the content. Surprisingly, however, when controlling for the predictability of the content, relation types that were more predictable led to longer reading times. We speculate that readers might want to confirm the relation type when their expectations about the content were incorrect, but leave this issue for further research.

Furthermore, we hypothesized that the facilitating effect of the connective would be proportional to the extent it reduces the predictability of the discourse relation. If so, variation in reading times in the presence vs. absence of a connective should be fully (or partially) attributed to differences in the unexpectedness of the relation type and content. We showed that the connective reduces the unexpectedness of the discourse relation type, and as a result also that of the content of the discourse relation. However, this did not significantly predict reading times. Instead, there was a strong facilitating effect of the connective, even when controlling for variation in the predictability of the relation type and content. Thus, we do not find evidence that the predictability of the discourse relation type and its content influences readers' sensitivity to connectives.

8.3 Characteristics of the reader

A third factor we examined is how characteristics of the reader might influence readers' sensitivity to discourse relation signals, in particular domain knowledge. As discussed in Chapter 2 (Section 2.5), discourse relation inference consists of understanding how the sentences are related (propositional integration) and relating this to existing knowledge (world-knowledge inference). When no linguistic signal is present, world-knowledge inference is needed to integrate the propositions. In the connective insertion study in Chapter 7, biomedical experts interpreted discourse relations in biomedical research papers more accurately than economics experts. This shows that domain knowledge helps readers to infer discourse relations. For economic texts, such an effect of domain knowledge was not found, possibly because the texts used in this study were accessible to a broader audience. In addition, readers resort to more underspecified interpretations when they were not an expert on the domain of the text, as we found more ambiguous connective insertions by non-experts. Thus, domain knowledge influences readers' inferences about discourse relations. We now turn to how domain knowledge influences readers' sensitivity to linguistic signals. In Section 8.3.2, we review other reader characteristics that have been discussed in previous literature.

8.3.1 The role of domain knowledge in exploiting discourse signals

The presence of linguistic cues could facilitate relational inference. We show that without linguistic cues, domain knowledge is indeed required to infer the relation, leading to more accurate interpretations by experts. We hypothesized that low-knowledge readers would exploit linguistic signals of the discourse relation to compensate for their lack of domain knowledge. Discourse signals were assumed to be more frequent in originally implicit relations, since readers cannot rely on connectives in these cases (cf. Sporleder & Lascarides, 2008; Hoek et al., 2019b; Crible, 2020). However, insertions in implicit relations were not significantly more accurate than those in implicitated relations. A qualitative analysis revealed that relation signals were present in relations that were originally signaled by a connective as well as those that were not. We also show that low-knowledge readers did not always exploit these cues. Partly, this could be attributed to the fact that relational cues sometimes require domain knowledge. For example, antonymy can signal `contrast` relations, but in order for a reader to recognize this, they need to know that there are linguistic elements that refer

to opposite concepts. Surprisingly, experts and non-experts also diverged when these signals could be recognized using general world knowledge. One possible explanation for this is that readers use non-connective cues to confirm their interpretation, rather than to explore different relation senses. In other words, expert readers might already consider a particular relation (e.g. `contrast`) based on their domain knowledge, for which the signal (e.g. antonymy) then provides additional evidence. Low-knowledge readers, on the other hand, lack the required domain knowledge to infer the relation and, as a result, can also not use the linguistic cue to confirm that interpretation.

Our findings extend earlier work on experts' and non-experts' use of connectives and other cohesive devices in comprehension. These studies show that low-knowledge readers benefit more strongly from the presence of connectives than high-knowledge readers (McNamara et al., 1996; McNamara, 2001; O'Reilly & McNamara, 2007; Kamalski et al., 2008). Thus, low-knowledge readers are sensitive to some linguistic signals of discourse relations. The cues manipulated in these earlier studies, however, are more informative: they signal a specific relation. Non-connective cues, on the other hand, are less informative in that they can occur in a wide variety of relations. Additional background knowledge is necessary in these cases to pinpoint the specific relation. The role of readers' background knowledge in their sensitivity to discourse cues thus interacts with the informativity of the signal.

8.3.2 Other reader characteristics

Other work has also shown that readers vary in their sensitivity to linguistic cues. For instance, readers differ in their interpretation of discourse relations (Scholman, 2019), how well they comprehend connectives, (e.g. Zufferey & Gygax, 2020b; Scholman et al., 2024a; Tskhovrebova et al., 2022a,b), to what extent non-connective signals modulate expectations about upcoming discourse relations (Scholman et al., 2020) and comprehenders' use of prosodic cues in distinguishing `subjective` and `objective` relations (Hu et al., 2023). Here, we discuss three reader characteristics that have been shown to affect the role of discourse relation signals in processing and representation.

The first factor that modulates readers' sensitivity to discourse cues is age. In order to effectively use a discourse cue to infer the relation, readers need to understand its meaning. The presence of a connective already influences on-line processing with children as young as eight years old (Cain & Nash, 2011), but readers' comprehension of connectives develops into adulthood (Nippold et al., 1992; Cain et al., 2005; Cain & Nash, 2011; Tskhovrebova et al., 2022a,b). Even adults do not show at ceiling performance in connective comprehension tasks (e.g. Zufferey & Gygax, 2020b; Scholman

et al., 2024b). This variation within adults can partly be attributed to characteristics of the signal such as lexical transparency and frequency (cf. Section 8.1.3), as well as to other reader-related differences.

One of these reader-related differences that influences connective comprehension is linguistic ability. Readers have been shown to struggle with the correct usage of connectives in a second language (Wetzel et al., 2020, 2023). Linguistic competence also influences connective comprehension in the native language. For instance, a larger vocabulary size predicts better understanding of connectives (Tskhovrebova et al., 2022a; Scholman et al., 2024a). In addition, participants who read more, as measured by an author recognition test (ART), have been found to show better performance on connective comprehension tasks (Zufferey & Gygax, 2020b; Tskhovrebova et al., 2022b; Wetzel et al., 2021, but see Tskhovrebova et al., 2022a; Scholman et al., 2024a). This factor could also partly explain age effects, as older readers will have gained more general linguistic experience. Furthermore, individual differences in linguistic ability also affect readers' sensitivity to discourse cues beyond connective comprehension. Scholman et al. (2020) show that print exposure modulates the extent to which readers use non-connective signals (quantifiers) to guide their expectations about upcoming `list` relations (but see Tskhovrebova et al., 2023, who do not find such an effect for French teenagers). This is in line with a statistical learing account of discourse signals: the more often a reader encounters the signal ∼ meaning mapping, the more likely they are to be sensitive to it. This could in part also explain the age effect, since older participants will have gained more linguistic experience over the course of their life. Thus, readers' sensitivity to discourse cues is influenced by their linguistic competence in various aspects of discourse representation.

A third factor that has been shown to influence to what extent readers (can) rely on linguistic signals for discourse relations is general intelligence. Tskhovrebova et al. (2022a) show that participants with higher academic level show better comprehension of connectives, across different age groups. Similarly, Scholman et al. (2024a) find that non-verbal IQ, as measured with Raven's Progressive Matrices Test (RPMT), also predicts the understanding of connectives. However, it is still unclear to what extent these factors influence the role of discourse cues in other aspects of representation and processing.[1]

[1] An exploratory analysis of the RPMT and ART data collected in Experiment 2 in Chapter 5 showed no effect of these measures on general reading speed, nor on the effect of the connective.

8.4 Characteristics of the language

In Chapter 4, we examined a fourth factor that might influence readers' sensitivity to discourse signals: characteristics of the language. Specifically, we hypothesized that the analyticity of the language would modulate the effect of the connective on subsequent processing. In analytic languages, information needs to be inferred from the context more often than in synthetic languages (cf. Blumenthal-Dramé, 2021). As a result, speakers of analytic languages might rely less on connectives when inferring discourse relations than speakers of synthetic languages. We conducted two self-paced reading experiments comparing the effect of the connective on reading in English (analytic) and German (synthetic). In neither of these experiments did we find evidence that the effect of the connective on processing is language-dependent. This suggests that whether a language is analytic or synthetic has a limited effect on the influence of the connective.

There might be other language-related factors that influence the role of discourse cues in processing and representation. For example, languages differ in how discourse relations are expressed. Many languages have connectives to distinguish `subjective` and `objective` relations, but this is not the case in English (cf. Sanders & Evers-Vermeul, 2019). However, in English, subjectivity can be expressed using prosody. In fact, there seems to be a trade-off between the use of lexical and prosodic cues in signaling discourse relations (Hu et al., 2022): Possibly, comprehenders are more sensitive to such prosodic cues in languages in which there are no lexical alternatives than in languages in which there are.

Furthermore, a statistical learning account of acquiring discourse signal meaning predicts that language-related differences in the distribution of discourse cues should influence the processing and representation of such cues. There is initial evidence that cross-linguistic variation in how relations are encoded affects processing. Yi & Koenig (2021) show that `cause` relations are usually not signaled by a connective in Korean. They find that Korean speakers are less likely to produce such relations compared to English speakers. This shows that there are language-related differences in expectations for discourse relations.

8.5 Conclusion

The research presented in this dissertation aimed to investigate which factors influence readers' sensitivity to signals of discourse relations. A summary of the findings

in Chapters 3 through 7 shows that the extent to which readers rely on discourse signals depends on characteristics of the signal itself, the relation and the reader. We do not find evidence that readers' sensitivity is modulated by the characteristic of the language we investigated, its analyticity. These findings were discussed together with related work, outlining other characteristics of the discourse signal, reader and language that influence to what extent comprehenders rely on discourse signals. In the next section, we will discuss suggestions for further research.

Chapter 9

Conclusion and outlook

9.1 Directions for future research

The research presented in this dissertation opens several directions for further research, both with respect to readers' sensitivity to discourse signals as well as more general theories of language processing.

9.1.1 Features of discourse relation signals

We proposed various features that seem to influence the role of the discourse signal in inferring the discourse relation (see Chapter 3). However, as discussed in Chapter 8, discourse cues often differ in multiple of these features. It is therefore still unclear whether all of these features *independently* influence the processing and representation of discourse relations. Further research is necessary to tease the effect of these features apart. This would require comparing the influence of discourse signals that differ only with respect to one feature, for example in an eye-tracking or self-paced reading experiment. Furthermore, we discussed that part of the reason why some features may or may not influence discourse processing and representation is their salience. Immutable lexical discourse cues whose primary function is to signal the relation are hypothesized to be more salient than non-lexical, non-specialized and mutable signals. To test this hypothesis, future research could examine whether highlighting discourse signals (e.g. using typographical and prosodic cues) increases readers' sensitivity to them.

Discourse signals need to provide information about the discourse relation. In Chapter 6, we argue that readers track statistical correlations between linguistic elements and discourse relations. However, the effect of this co-occurrence on processing seems to be limited. This raises the question what amount of statistical information the signal needs to provide in order to facilitate discourse relation inference. In other words, how often does the reader need to be exposed to the co-occurrence of a cue with a discourse relation, in order for them to pick this up as a signal? This could be examined by comparing the effect of cues that vary in their informativity. The informativity of these cues could either be based on general occurrences in a corpus or could also be manipulated by exposing readers to certain signal $\sim$ meaning mappings before testing (cf. Johnson & Arnold, 2023, for a similar framework for investigating referential biases).

The features proposed in this dissertation are likely not exhaustive: there might be other signal characteristics that influence whether readers are sensitive to the signal. One such feature that has not been discussed in this dissertation is whether the signal is in an at-issue clause. For example, Yao et al. (2024) show that implicit causality verbs in at-issue clauses raise stronger questions-under-discussion (cf. Clifton Jr & Frazier, 2012; Kehler & Rohde, 2017) about causal relations than when these verbs are in clauses that are not at-issue. Further research is necessary to investigate whether this also applies to other types of signals and clauses.

Connectives have been shown to elicit two distinct processes: propositional integration and world-knowledge inference (cf. Cozijn et al., 2011). It is still unclear if the same processes are triggered by non-connective cues. After all, non-connective signals are not as informative as connectives, so readers cannot only rely on the signal to determine the relation. Instead, readers might still need their world-knowledge to infer the discourse relation. In Chapter 7, we hypothesize that readers possibly use non-connective cues only to confirm the relation type that they established based on existing knowledge. This suggests that the content of the relation has a strong effect on the interpretation of non-connective cues. Previous studies indeed find that effects of non-connective cues show up strongly in off-line continuation tasks (cf. Kehler et al., 2008; Scholman et al., 2020, Chapter 6), where no content is provided, but such findings are mixed for tasks on readers' sensitivity to such cues when it can also be derived based on the content (cf. Crible & Pickering, 2020, Chapters 6 and 7). However, further research is needed to investigate whether and how the processes that are elicited by connectives and non-connective cues differ. For example, do non-connective signals for causal relations also facilitate sentence-initial, but not

sentence-final reading (cf. Cozijn et al., 2011; Millis & Just, 1994)? And are readers indeed less sensitive to non-connective signals when they have no clue about the relation sense compared to when they need to disambiguate between only a few relation types?

9.1.2 Individual differences in discourse processing

Readers differ in how they interpret discourse relations. In Chapter 7, we show that individual differences in relation interpretation can also be attributed to the readers' background knowledge. We argue that readers resort to underspecified interpretations in the absence of the required domain knowledge. This raises a number of questions for further research about what such underspecified interpretations look like, whether they are task-dependent (cf. Noordman et al., 1992) and whether they are also influenced by other reader characteristics (e.g. processing depth, cf. Scholman, 2019, Chapter 8). Furthermore, readers may have biases towards certain interpretations. Sanders (2005) argues that readers have a preference for causal interpretations. Although we did not find evidence that readers resort to such default interpretations in the absence of domain knowledge (see Chapter 7), Scholman (2019, Chapter 8) shows that some, but not all, readers prefer argumentative interpretations of `specification` relations. It would be interesting to see what reader-related characteristics influence discourse relation interpretations.

Furthermore, differences between readers may also affect their sensitivity to linguistic signals of the discourse relation. Age, linguistic proficiency and general intelligence have been shown to influence how well readers comprehend connectives (e.g. Tskhovrebova et al., 2022b; Scholman et al., 2024a), but it is unclear whether there is also individual variability in the extent to which connectives influence on-line processing. This would first require establishing whether individual differences between participants in the effect of the connective (e.g. in a similar self-paced reading or eye-tracking-while-reading setup as in Chapters 4 or 5) are reliable (e.g. stable across different sessions, cf. Staub, 2021; Haller et al., 2023; Frinsel & Christiansen, 2024). If so, one could examine whether variability in this effect is explained by factors that have been shown to affect connective comprehension or on-line language processing (e.g. working memory, Nicenboim et al., 2015).

In addition, previous research on individual differences in discourse processing has mostly focused on connectives. Chapter 7 reveals that domain knowledge is sometimes required to interpret non-connective cues, suggesting that there are individual differences in the comprehension of other signals as well. Further research is needed

to examine whether the factors that influence connective comprehension also explain individual variability in readers' sensitivity to non-connective signals. If the meaning of discourse signals is acquired by their statistical co-occurrence with the discourse relation, as argued in Chapter 6, linguistic experience likely also influences readers' sensitivity to other non-linguistic cues. Indeed, Scholman et al. (2020) show that readers with more reading experience provide more `list` continuations after a quantifier in the context. However, more research is needed to examine whether this also extends to other discourse signals, such as non-lexical cues, as well as to other tasks, such as on-line processing studies. Finally, the lower salience and informativity of non-connective cues might make them more difficult to exploit, which might magnify differences between readers. Future research could examine whether individual differences play a more important role in the sensitivity to non-connective cues compared to connectives.

9.1.3 The role of prediction in discourse processing

Comprehenders continually make predictions about upcoming material. The unexpectedness of linguistic material (Levy, 2008) has been argued to be proportional to processing difficulty. However, it is unclear which aspects of that material readers make predictions about and to what extent they are guided by bottom-up input. In Chapter 5, we investigated whether the processing of discourse relations can be explained by the predictability of the discourse relation type and the content. Although predictable content was read faster across different measures, the results were less clear with respect to the predictability of the relation type. The finding that relations are read slower when the relation type is more predictable is surprising and requires further replication. In addition, we speculated that readers might want to confirm their prediction about the relation type even when the content is different from what they expected, but further research is necessary to confirm this. Evidence for this hypothesis could be found by examining whether the effect of content predictability is stronger in high-constraint compared to low-constraint conditions.

Furthermore, predictability effects cannot fully account for differences in processing difficulty. In Chapter 5, we found additional effects of the presence of a connective when controlling for predictability effects (cf. Huang et al., 2024, for a similar study on syntactic disambiguation). One explanation for such a finding is that our measures of unexpectedness are not adequate. Further research is needed to examine how the predictability of linguistic material can be captured accurately. Another explanation is that there are other mechanisms beyond prediction that influence processing dif-

ficulty (cf. Huang et al., 2024). In Chapter 5, we suggest that readers may adapt their processing strategy in the presence of a connective. Future work should test what these processes are. For example, do readers engage in more shallow processing following the connective? How does this influence subsequent representation?

9.1.4 Cross-linguistic research on discourse processing

Finally, we encourage further work on cross-linguistic differences in discourse processing. In Chapter 4, we found no evidence that the facilitating effect of the connective depends on the typology of the language. However, this does not mean that there is no cross-linguistic variation in readers' sensitivity to discourse signals. We investigated two closely related languages, but such differences might be more pronounced in languages that are on more extreme ends of the analytic-synthetic continuum, such as Mandarin Chinese and Russian. Furthermore, language-related differences in the occurrence of discourse relations or their signals may influence readers' processing of these relations (cf. Yi & Koenig, 2021). Future work could examine how languages differ in their distribution of discourse relation types (e.g. using corpora) and how such statistical information influences the processing of discourse relations. The latter can be examined by comparing cues that carry a similar meaning across languages, but differ in their informativity across languages.

In addition, future work should focus on how discourse relations are signaled in other languages. This could provide more insights into the type of cues, other than connectives, that could signal discourse relations. For example, verbs or syntactic structure might be more common signals of discourse relations in languages other than English (cf. Zeyrek et al., 2020; Marchal et al., 2021b). Comprehenders of these languages might be more sensitive to such non-connective cues than speakers of English.

One area of cross-linguistic research that is especially interesting are contact languages. These languages are still developing and speakers might not always be able to express discourse relations with a connective, simply because such a connective does not (yet) exist. This raises various questions for further research: How are discourse relations expressed in these languages? Which relations are expressed by a connective and which by other means? And do speakers of such languages show similar sensitivity to connectives than speakers of languages in which specialized signals for discourse relations are more common? These question can be answered by constructing discourse-annotated corpora for these languages and assessing whether speakers

of contact languages similarly rely on connectives in continuation tasks or reading studies.

9.2 Conclusion

The research presented in this dissertation sheds more light on how readers establish coherence, a requisite for successful comprehension. We show that readers exploit linguistic cues (i.e. connectives) when processing discourse relations, but also use extra-linguistic information (i.e. their domain knowledge). In addition, our findings reveal that the extent to which readers are sensitive to linguistic signals of discourse relations depends on characteristics of that signal (e.g. connective vs. non-connective cues), the relation type (e.g. `reason` vs. `concession`) and the reader (i.e. high vs. low-knowledge readers).

These findings contribute to theories on the processing of discourse relations, by highlighting the need for such theories to also include readers' sensitivity to linguistic signals other than connectives and account for when these signals are exploited. The results presented in this dissertation also provide insights that are important for information-theoretic accounts, by showing limits to the role of prediction in discourse processing, and theories of statistical learning, by demonstrating that readers also track signal∼meaning correlations at the discourse-level. Combining insights from various fields and methodologies can help us to construct better theories and opens new directions for research. Further research on how readers make sense of discourse will lead to a deeper understanding of how we are able to communicate so efficiently.

Appendices

A Chapter 2: PDTB-3 hierarchy

Table is printed on the next page.

Table 1: Hierarchy of the PDTB-3 labels (Webber et al., 2019).

Level-1	**Level-2**	**Level-3**
temporal	synchronous	
	asynchronous	precedence
		succession
contingency	cause	reason
		result
		negresult
	cause+belief	reason+belief
		result+belief
	cause+speechact	reason+speechact
		result+speechact
	condition	arg1-as-cond
		arg2-as-cond
	condition+speechact	
	negative-condition+speechact	arg1-as-negcond
		arg2-as-negcond
	negative-condition+speechact	
	purpose	arg1-as-goal
		arg2-as-goal
comparison	concession	arg1-as-denier
		arg2-as-denier
	concession+speechact	arg2-as-denier+speechact
	contrast	
	similarity	
expansion	conjunction	
	disjunction	
	equivalence	
	exception	arg1-as-excpt
		arg2-as-excpt
	instantiation	arg1-as-instance
		arg2-as-instance
	level-of-detail	arg1-as-detail
		arg2-as-detail
	manner	arg1-as-manner
		arg2-as-manner
	substitution	arg1-as-subst
		arg2-as-subst

B Chapters 3 & 8: Overview of evidence of relation signals influencing discourse relation representation and processing

An extended version of Table 3.2 in Chapter 3 (see also Table 8.1 in Chapter 9) is repeated here as Table 2. Evidence for the effects of each type of cue is provided below.

Table 2: Overview of example discourse signals with their features, as well as their effect on representation or processing. For the features, black bullets indicates that the feature is (strongly) present. Gray bullets show that the feature is less strong or absent. For the findings, black squares indicates that there is strong evidence for the effect of this type of discourse signal, with ■ indicating smaller effects or mixed findings and ■ representing that no effect has been found. -: not applicable

	because; nevertheless	*and; but; when*	*this caused; due to rain*	*several; true*	*not*; antonyms	implicit causality	tense; eventuality	parallelism	gerund free adjuncts
functionality	●	●	●	●	●	●	●	●	●
informativity	●	●	●	●	●	●	●	●	●
immutability	●	●	●	●	●	●	●	●	●
lexicality	●	●	●	●	●	●	●	●	●
agreement	-	-	-	●	●	●	●	●	●
representation	■[1]	■[2]	■[3]	■[4]	■[5]	■[6]	■[7]	■[8]	■[9]
processing	■[10]	■[11]	■[12]	■[13]	■[14]	■[15]	■[16]	■[17]	■[18]

■ **1** Relation-specific connectives facilitate comprehension (see e.g. Kleijn et al., 2019; Van Silfhout et al., 2014) and elicit off-line expectations about upcoming discourse relations (Scholman et al., 2017). Annotation agreement is higher on explicit compared to implicit relations (Hoek et al., 2021c).

■ **2** Readers use statistical information about the co-occurrence of ambiguous connectives to guide their interpretation (Asr & Demberg, 2020). Nevertheless, annotation agreement on relations marked by ambiguous connectives is lower than on relations marked by a relation-specific connective compared, but not higher than in implicit relations (Hoek et al., 2021c).

■ **3** Resultative verbs facilitate the interpretation of `result` relations (Crible & Demberg, 2020). However, Sanders & Noordman (2000) find no evidence that free connecting phrases facilitate recall.

■ **4** Quantifiers in the context elicit expectations about upcoming `list` relations (Scholman et al., 2020; Tskhovrebova et al., 2023), although these are less strong than found by connectives and there are individual differences in this effect (Scholman et al., 2020).

■ **5** Antonyms facilitate the interpretation of `contrast` relations (Crible & Demberg, 2020).

■ **6** Implicit causality verbs have repeatedly been shown to influence off-line expectations about upcoming discourse relations (Simner & Pickering, 2005; Kehler et al., 2008; Bott & Solstad, 2021).

■ **7** Eventuality type, state duration and event complexity has been shown to influence inferences about temporal order (Dery & Koenig, 2015; Marx et al., 2024).

■ **8** No effect of parallelism has been found for the speed or accuracy of verifying `contrast` relations (Crible & Pickering, 2020).

■ **9** Chapter 6 shows that readers' expectations for upcoming discourse relations are influenced by gerund free adjuncts, but we do not find evidence that clause structure affects readers' preferences for how the discourse relation is formulated.

■ **10** Relation-specific connectives lead to shorter reading times of the discourse relation (see e.g. Cozijn et al., 2011; Van Silfhout et al., 2014) and lead to on-line predictions of upcoming material (Köhne-Fuetterer et al., 2021), facilitating the processing of that material (Xiang & Kuperberg, 2015; Köhne-Fuetterer et al., 2021). Furthermore, inappropriate connectives disrupt processing (Murray, 1997; Canestrelli et al., 2013; Xu et al., 2018; Wetzel et al., 2021).

■ **11** Relations that are signaled by ambiguous connectives are processed faster than relations that are not signaled by a connective (Zufferey & Gygax, 2016), but slower than those signaled by relation-specific connectives (Cain & Nash, 2011; Li et al., 2017). Ambiguous connectives disrupt the processing of relations that they do not signal (Koornneef & Sanders, 2013; Wetzel et al., 2021), while allowing for interpretations that they can signal (Mak & Sanders, 2013).

■ **12** Free connecting phrases facilitate the processing of subsequent discourse relations compared to relations without such a signal (Sanders & Noordman, 2000).

■ **13** Lexical discourse signals, like *true*, have been found to facilitate the processing of an upcoming connective also signaling that relation, although this effect was not as strong across studies (Schwab & Liu, 2020).

■ **14** The presence of negation reduces the processing disadvantage of `concession` relations, but does not facilitate the processing of a `concession` connectives (Crible, 2021).

■ **15** Readers can make on-line predictions about upcoming discourse relations based on implicit causality verbs (Rohde & Horton, 2014) and these verbs also facilitate the processing of causal discourse relations (see e.g. Mak & Sanders, 2013; Hoek et al., 2021a).

■ **16** Temporal connectives, but not tense, has been found to lead to faster processing of chronological discourse relations (Grisot & Blochowiak, 2017).

■ **17** Parallelism facilitates the processing of `contrast` relations more strongly in the presence of an ambiguous than a relation-specific connective, but only when readers are asked to disambiguate the relation (Crible & Pickering, 2020).

■ **18** We find no evidence that gerund free adjuncts influence the processing of `result` relations in Chapter 6.

C Chapter 4, Experiment 2: word-by-word analysis

The results for our region-by-region analysis using the analysis approach explained in Section 4.5.1 can be found in Table 3 and in Figure 1. For the critical region, both language and the interaction between marking and relation are significant. In the spillover region, marking, relation, language and the interaction between marking and relation are significant. In the final region, marking and language are significant. In these last two regions, implicit marking slows down reading times. The German stimuli have slower reading times in all three regions. Note that the interaction between marking and language does not reach significance in any of the three regions.

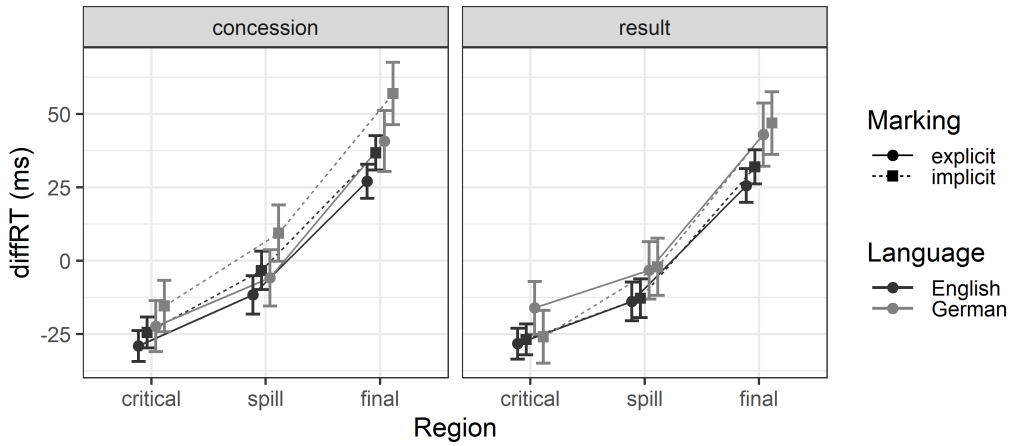

Figure 1: Fitted reading times of Experiment 2 per condition and per region (separate models).

Table 3: Model outputs for the word-by-word analysis in Experiment 2

	critical				spillover				final			
	β	SE	t	p	β	SE	t	p	β	SE	t	p
(Intercept)	-23.51	1.74	-13.50	**<.001**	-5.39	2.39	-2.26	**0.03**	38.67	1.87	20.64	**<.001**
marking	-0.38	1.14	-0.33	0.74	-3.23	1.03	-3.14	**<.01**	-4.52	1.42	-3.19	**<.01**
relation	-0.73	1.17	-0.63	0.53	-2.57	1.04	-2.48	**0.01**	-1.76	1.42	-1.24	0.22
language	-3.61	1.74	-2.07	**0.04**	-4.98	2.39	-2.08	**0.04**	-8.28	1.87	-4.42	**<.001**
marking:rel	2.49	1.11	2.24	**0.03**	2.65	1.03	2.58	**0.01**	1.96	1.42	1.38	0.17
marking:lang	-1.12	1.14	-0.99	0.32	0.88	1.03	0.86	0.39	0.51	1.42	0.36	0.72
relation:lang	0.34	1.17	0.29	0.77	-0.35	1.04	-0.34	0.73	0.20	1.42	0.14	0.89
mark:rel:lang	-1.69	1.11	-1.52	0.13	-0.85	1.03	-0.82	0.41	-1.12	1.42	-0.79	0.43

Model formula (critical): `diffRT ~ mark*rel*lang+(1+rel||item)+(0+mark||subj)`

Model formula (spill): `diffRT ~ mark*rel*lang+(1+rel||item)`

Model formula (final): `diffRT ~ mark*rel*lang+(1|item)`

D Chapter 5: Additional pretests

D.1 Pretest 1: Predictability

To examine the effects of predictability, items were needed that vary with respect to both relational and semantic surprisal. This first pretest was therefore run to select materials that would differ in the predictability of the content of the target region in the presence of a connective, since we expected that this would be the most difficult manipulation of the materials. The goal of this first pretest was to obtain items in which the target region should be highly predictable from the context in one condition, but not in the other. Note that these conditions were conceptualized solely with the purpose of creating variance in the predictability in our items. In the studies presented in Chapter 5, we used item- and context-specific continuous measures of predictability.

The materials consisted of 40 prompts (containing a connective to ensure `result` continuations) in two conditions. As can be seen in (64), the items in the two conditions differed in the context that was presented to the participants, but had the same target region, which was not shown. This context was assumed to lead to continuations that were highly predictable (`HP`) or less predictable (`LP`).[1]

(64) Angela used to live in a small flat in Atlanta. **target: was evicted**

a. She didn't pay rent for months. Therefore, she ... `HP`

b. She had over fifteen cats. Therefore, she ... `LP`

This target region was assumed to be predictable in one condition, but not in the other. To test this assumption, participants (n = 46) were asked to fill out logical and plausible continuations to the prompt. The items were distributed across 4 lists, with every participant seeing 10 items in each condition. This resulted in 11-12 observations per item per condition. The continuations were blindly annotated for whether they were (a paraphrase of) the target region or not.[2] Continuations for the example item above with their annotations can be found below.

(65) `target: was evicted`

a. ... was evicted. `target`

b. ... was kicked out. `target`

[1] Note that for some items, the first sentence also differed across conditions.

[2] Agreement on the first round of this pretest was 82%.

c. ... finally bought a house. `non-target`

We then selected those items in which the target region was predicted in at least 60% of cases (i.e. by 7/11 or 8/12 participants) in the `HP` condition, and had a cloze probability of at least 40% lower in the `LP` condition. Since not enough items (n=23) met these requirements after a first pretest, some items (n=18) were improved and retested in a second round, with 24 different participants. After this second round, 31 items satisfied the requirements described above and were included in the second pretest, which examined the plausability of the items. The table below shows a summary of the cloze probability of the final set of items included in the self-paced reading experiment.

Table 4: Summary of cloze probabilities of final item set

	mean	**min**	**max**
`HP`	0.78	0.64	1.00
`LP`	0.13	0.00	0.36
difference	0.65	0.42	1.00

D.2 Pretest 2: Plausibility

The second pretest tested if all items were considered plausible in all conditions. The materials consisted of the 31 items selected from the previous pretest, but now also the target region and the spill-over region were presented to the participants. Each item was presented in 2 (with vs. without connective) x 2 (high vs. low predictability).

(66) Angela used to live in a small flat in Atlanta.

a. She didn't pay rent for months. Therefore, she was evicted by her landlord.
`HP | exp`

b. She didn't pay rent for months. She was evicted by her landlord.
`HP | imp`

c. She had over fifteen cats. Therefore, she was evicted by her landlord.
`LP | exp`

d. She had over fifteen cats. She was evicted by her landlord.
`LP | imp`

The items were distributed across four lists, and dispersed with ten implausible, but grammatical fillers, illustrated in (68).

(67) Sally was meeting with her son Ike's school counselor. Ike had beaten up a classmate. Therefore, he was commended by the school.

(68) Owen was throwing a birthday party and had invited everyone he knew. Lucy had been feeling nauseous all day. Therefore, she went to see a doctor.

Participants (n=95) were asked to rate how plausible they thought each item was on a scale of 1-5 stars. Every participant saw 3-4 items in each condition, resulting in 11-14 observations per item per condition. Overall, fillers (mean = 2.42) were rated less plausible than experimental items (mean = 4.26). We selected those 24 items with a mean of at least 3 in all conditions. In addition, we only included items for which the mean plausibility did not differ too much (< 1) across conditions. The table below shows a summary of the plausibility ratings of the final set of items included in the self-paced reading experiment.

Table 5: Mean plausibility rating of final item set

	exp	imp
HP	4.48	4.47
LP	4.09	4.24

E Chapter 6: Additional information on corpus methodology

E.1 PDTB

Entity relations as well as explicit relations were excluded from the selected adjuncts. In the PDTB, the free adjunct is always annotated as the second argument, regardless of its position in the sentence. Since the first clause in the text is considered to be the first argument for implicit relations, we recoded the relation senses, depending on the position of the free adjunct compared to the main clause. To illustrate, a `arg2-as-detail` relation would be recoded as `arg1-as-detail` if the GFA was sentence-final, since the first argument in the text provides the detail. We automatically determined the textual position of the GFA by examining whether the second argument occurred before (initial), within (medial) or after the first argument (final). Since we did not make a distinction between the features `+Belief` and `+SpeechAct`, we combined these with their main relation sense. Furthermore, our definition of alternative lexicalizations diverges from the one taken in the PDTB, where alternative lexicalizations are only identified when inserting a connective is redundant. However, these heuristics do not always make it clear in which instances the gerund itself should be considered an alternative lexicalization and could therefore not be applied systematically. For example, in the PDTB, GFA's headed by *killing* was considered an alternative lexicalization, whereas *using* was not. This was reversed in our annotations, since *using* can often be replaced with *by*, whereas there is a loss of meaning when *killing* is replaced by a causal connective. All cases in the PDTB were therefore revised to ensure consistency in the annotations of alternative lexicalizations. To examine the proportion of `result` relations in non-GFA's, we consider the relational distribution of implicit inter-sentential relations in the PDTB3. Since inter-sententiality is not provided, we selected those instances where both arguments start with a capital letter. In addition, we included those instances where a PB-role was indicated. A check of 100 random instances in the PDTB. We identified 12402 instances in the PDTB. Note that this is lower than what is written in the PDTB3 manual, possibly due to our more coarse-grained method.

E.2 BAC

Gerund free adjuncts form the Blog Authorship Corpus were sourced by parsing the first 4,750 files of the corpus using Spacy. We then selected `adverbial clause modifiers (advcl)` and `open clausal complements (xcomp)` ending in *-ing*. To refine this automatic search and exclude present participles that are verbal complements instead of free adjuncts and excluded hits that modified verbs of emotion and attitude (*enjoy laughing*), perception (*watch him laughing*), as well as existential (*sit laughing)* and phase-verbs (*started laughing*) and aspectualizers (*keep laughing*) (Kortmann, 2013). In addition, we excluded GFAs that are explicitly marked by a connective (e.g. *While having a beer, John talked to his friends.*), by only selecting these instances in which the gerund candidate was the left-most element in the clause. Since the text style in the BAC varies widely and is not always grammatical, the automatic identification of the GFA needed to be manually verified. We excluded instances that could not be moved to another position in the sentence without changing their syntactic position. as in the example in 72. This analysis distinguishes whether the free adjunct is attached to the entire matrix clause or its subject on the one hand, in which case it is included in the analysis, compared to when it is attached to the object of the matrix clause (as in 70).

(69) a. [Assuming we could get people to land in New Orleans at close to the same times (or a couple of times)], I'm not opposed to providing transportation for the weekend.

b. I'm not opposed to providing transportation for the weekend, [assuming we could get people to land in New Orleans at close to the same times (or a couple of times)].

(70) A picture taken at that time shows the village in and around their first bus and her young father on his bicycle [smiling].

Note that the restrictions on the change in meaning are purely syntactic, not discourse-structural, since the relation sense often changes when moving the position of the free adjunct. This is illustrated in Example 72, where the original relation sense is `precedence`, but `arg1-as-manner` when the free adjunct is moved to the initial position.

(71) Get back to 95, going South this time to Petersburg.

(72) Going South this time to Petersburg, get back to 95.

The textual position of the gerund free adjunct with respect to the main clause was identified manually in this corpus.

F Chapter 7: Relation sense classification

The connectives in the second step were categorized as signaling eight different relational classes: (1) cause, (2) temporal, (3) contrast, (4) concession, (5) positive expansion (e.g. `instantiation`), (6) negative expansion (e.g. `disjunction`), (7) condition, (8) no relation. These classes included connectives signaling the following PDTB3 level 2 relation senses:

1. Cause: `cause`, `cause+belief`, `cause+speechact`, `purpose`
2. Temporal: `synchronous`, `asynchronous`
3. Contrast
4. Concession: `concession`, `concession+speechact`
5. Positive expansion: `similarity`, `conjunction`, `equivalence`, `instantiation`, `level-of-detail`, `manner`
6. Negative expansion: `disjunction`, `exception`, `substitution`
7. Condition: `condition`, `condition+speechact`, `negative-condition`, `negative-condition+speechact`

G Chapter 7: Means across conditions

Table 6: Mean percentage of correct answers per condition.

		BioDRB		**PDTB**		
		bio	eco	bio	eco	*mean*
Implicitated	Result	57.3	47.2	72.6	72.1	*62.4*
	Instantiation	71.4	41.0	46.8	49.5	*52.0*
	Concession	48.4	28.6	48.9	58.6	*48.6*
	Contrast	34.7	27.6	42.9	14.3	*31.2*
	mean	*52.2*	*36.2*	*54.3*	*58.7*	
Implicit	Result	70.3	69.1	64.1	65.1	*67.1*
	Instantiation	65.1	51.9	64.1	60.5	*60.7*
	Concession	41.0	36.1	54.9	54.2	*49.1*
	Contrast	39.1	25.9	33.3	20.0	*31.7*
	mean	*54.1*	*45.4*	*58.7*	*57.0*	

List of Figures

List of Tables

Bibliography

Altmann, G., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. doi:10.1016/S0010-0277(99)00059-1.

Angelo Canty, & B. D. Ripley (2024). *boot: Bootstrap R (S-Plus) Functions*. R package version 1.3-30.

Arnold, J. E., Strangmann, I. M., Hwang, H., Zerkle, S., & Nappa, R. (2018). Linguistic experience affects pronoun interpretation. *Journal of Memory and Language*, *102*, 41–54. doi:10.1016/j.jml.2018.05.002.

Artstein, R., & Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics*, *34*(4), 555–596. doi:10.1162/coli.07-034-R2.

Asher, N., & Lascarides, A. (2003). *Logics of conversation*. Cambridge University Press.

Asr, F. T., & Demberg, V. (2012). Implicitness of discourse relations. In *Proceedings of the International Conference on Computational Linguistics (COLING)* (pp. 2669–2684). Mumbai, India.

Asr, F. T., & Demberg, V. (2013). On the information conveyed by discourse markers. In *Proceedings of the Fourth Annual Workshop on Cognitive Modeling and Computational Linguistics (CMCL)* (pp. 84–93).

Asr, F. T., & Demberg, V. (2015). Uniform Information Density at the level of discourse relations: Negation markers and discourse connective omission. In *Proceedings of the International Conference on Computational Semantics (IWCS)* (pp. 118–128). London, UK.

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. doi:10.1080/0163853X.2019.1700760.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. doi:10.1016/j.jml.2007.12.005.

Bachand, F.-H., Davoodi, E., & Kosseim, L. (2014). An investigation on the influence of genres and textual organisation on the use of discourse relations. In *International Conference on Intelligent Text Processing and Computational Linguistics* (pp. 454–468). Springer.

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. doi:10.1016/j.jml.2012.11.001.

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

Bates, D., & Sarkar, D. (2007). The lme4 package. *R package version*, *2*(1).

Becher, V. (2011). *Explicitation and implicitation in translation*. Ph.D. thesis, Universität Hamburg.

Blochowiak, J., Grisot, C., & Degand, L. (2022). From implicit to explicit: The processing of forward causal and temporal relations. *Pragmatics & Cognition*, *29*(1), 29–58. doi:10.1075/pc.21001.blo.

Blumenthal-Dramé, A. (2021). The online processing of causal and concessive relations: comparing native speakers of English and German. *Discourse Processes*, *58*(7), 642–661. doi:10.1080/0163853X.2020.1855693. Retraction published 29 Apr 2024, Discourse Processes 61(3), 164. `https://www.tandfonline.com/doi/full/10.1080/0163853X.2024.2348297`.

Bott, O., & Solstad, T. (2021). Discourse expectations: Explaining the implicit causality biases of verbs. *Linguistics*, *59*(2), 361–416. doi:10.1515/ling-2021-0007.

Bouma, G. (2009). Normalized (pointwise) mutual information in collocation extraction. (pp. 31–40). Potsdam volume 30.

Bourgonje, P., Hoek, J., Evers-Vermeul, J., Redeker, G., Sanders, T. J., & Stede, M. (2018). Constructing a lexicon of Dutch discourse connectives. *Computational Linguistics in the Netherlands Journal*, *8*, 163–175.

Boyce, V., Futrell, R., & Levy, R. P. (2020). Maze made easy: Better and easier measurement of incremental processing difficulty. *Journal of Memory and Language*, *111*, 104082. doi:10.1016/j.jml.2019.104082.

Britton, J. R., Cong, Y., Hsu, Y.-Y., Chersoni, E., & Blache, P. (2024). On the influence of discourse connectives on the predictions of humans and language models. *Frontiers in Human Neuroscience*, *18*, 1–17. doi:10.3389/fnhum.2024.1363120.

Van den Broek, P. (2010). Using texts in science education: Cognitive processes and knowledge representation. *Science*, *328*, 453–456. doi:10.1126/science.1182594.

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. doi: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. doi: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. doi:10.1080/01690965.2012.685885.

Carlson, L., & Marcu, D. (2001). Discourse tagging reference manual. *ISI Technical Report ISI-TR-545*, *54*, 1–56.

Carroll, L. (1993). *Alice in Wonderland and Through the Looking Glass*. Wordsworth Editions.

Chen, L., Paterson, K. B., Li, X., Li, L., & Yang, Y. (2019). Pragmatic influences on sentence integration: Evidence from eye movements. *Quarterly Journal of Experimental Psychology*, *72*(12), 2742–2751. doi:10.1080/01690965.2012.685885.

Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. Van Gompel, M. Fischer, W. Murray, & R. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 341–371). Oxford: Elsevier.

Clifton Jr, C., & Frazier, L. (2012). Discourse integration guided by the 'question under discussion'. *Cognitive Psychology*, *65*(2), 352–379. doi:10.1016/j.cogpsych.2012.04.001.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46. doi:10.1177/001316446002000104.

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, *114*(3), 356–371. doi:10.1016/j.cognition.2009.10.009.

Couper-Kuhlen, E. (1996). Intonation and clause combining in discourse: The case of because. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)*, *6*(3), 389–426. doi:10.1075/prag.6.3.04cou.

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. doi:10.1080/0163853X.2011.594421.

Crible, L. (2020). Weak and strong discourse markers in speech, chat, and writing: Do signals compensate for ambiguity in explicit relations? *Discourse Processes*, *57*(9), 793–807. doi:10.1080/0163853X.2020.1786778.

Crible, L. (2021). Negation cancels discourse-level processing differences: Evidence from reading times in concession and result relations. *Journal of Psycholinguistic Research*, *50*(6), 1–26. doi:10.1007/s10936-021-09802-2.

Crible, L. (2022). The syntax and semantics of coherence relations: From relative configurations to predictive signals. *International Journal of Corpus Linguistics*, *27*(1), 59–92. doi:10.1075/ijcl.19109.cri.

Crible, L., & Degand, L. (2019). Reliability vs. granularity in discourse annotation: What is the trade-off? *Corpus Linguistics and Linguistic Theory*, *15*(1), 71–99. doi:10.1515/cllt-2016-0046.

Crible, L., & Demberg, V. (2020). The role of non-connective discourse cues and their interaction with connectives. *Pragmatics & Cognition*, *27*(2), 313–338. doi:10.1075/pc.20003.cri.

Crible, L., & Pickering, M. J. (2020). Compensating for processing difficulty in discourse: Effect of parallelism in contrastive relations. *Discourse Processes*, *57*(10), 862–879. doi:10.1080/0163853X.2020.1813493.

Crosson, A. C., & Lesaux, N. K. (2013). Does knowledge of connectives play a unique role in the reading comprehension of English learners and English-only students? *Journal of Research in Reading*, *36*(3), 241–260. doi:10.1111/j.1467-9817.2011.01501.x.

Dahl, D. B., Scott, D., Roosen, C., Magnusson, A., & Swinton, J. (2019). *xtable: Export Tables to LaTeX or HTML*. R package version 1.8-4.

Dale, R. (1991). Exploring the role of punctuation in the signalling of discourse structure. In *Proceedings of a Workshop on Text Representation and Domain Modelling: Ideas from Linguistics and AI* (pp. 110–120).

Danlos, L., Rysova, K., Rysova, M., & Stede, M. (2018). Primary and secondary discourse connectives: Definitions and lexicons. *Dialogue & Discourse*, *9*(1), 50–78. doi:10.5087/dad.2018.102.

Das, D., Scheffler, T., Bourgonje, P., & Stede, M. (2018). Constructing a lexicon of English discourse connectives. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue* (pp. 360–365).

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).

Das, D., & Taboada, M. (2018a). RST Signalling Corpus: A corpus of signals of coherence relations. *Language Resources and Evaluation*, *52*(1), 149–184. doi:10.1007/s10579-017-9383-x.

Das, D., & Taboada, M. (2018b). Signalling of coherence relations in discourse, beyond discourse markers. *Discourse Processes*, *55*(8), 743–770. doi:10.1080/0163853X.2017.1379327.

De Varda, A. G., Marelli, M., & Amenta, S. (2023). Cloze probability, predictability ratings, and computational estimates for 205 English sentences, aligned with existing EEG and reading time data. *Behavior Research Methods*, *56*, 5190—-5213. doi:10.3758/s13428-023-02261-8.

Degand, L., & Sanders, T. J. (2002). The impact of relational markers on expository text comprehension in L1 and L2. *Reading and writing*, *15*, 739–757. doi:10.1023/A:1020932715838.

Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, *109*(2), 193–210. doi:10.1016/j.cognition.2008.07.008.

Demberg, V., Kravtchenko, E., & Loy, J. E. (2023). A systematic evaluation of factors affecting referring expression choice in passage completion tasks. *Journal of Memory and Language*, *130*, 104413. doi:10.1016/j.jml.2023.104413.

Demberg, V., Scholman, M. C., & Asr, F. T. (2019). How compatible are our discourse annotation frameworks? Insights from mapping RST-DT and PDTB annotations. *Dialogue & Discourse*, *10*(1), 87–135. doi:10.5087/dad.2019.104.

Den Ouden, H., Noordman, L., & Terken, J. (2009). Prosodic realizations of global and local structure and rhetorical relations in read aloud news reports. *Speech Communication*, *51*(2), 116–129. doi:10.1016/j.specom.2008.06.003.

Dery, J. E., & Koenig, J.-P. (2015). A narrative-expectation-based approach to temporal update in discourse comprehension. *Discourse Processes*, *52*(7), 559–584. doi:10.1080/0163853X.2014.966293.

Duque, E. (2014). Signaling causal coherence relations. *Discourse Studies*, *16*(1), 25–46. doi:10.1177/1461445613496358.

Enochson, K., & Culbertson, J. (2015). Collecting psycholinguistic response time data using amazon mechanical turk. *PLOS One*, *10*(3), e0116946. doi:10.1371/journal.pone.0116946.

Evers-Vermeul, J., Hoek, J., & Scholman, M. C. (2017). On temporality in discourse annotation: Theoretical and practical considerations. *Dialogue & Discourse*, *8*(2), 1–20. doi:10.5087/dad.2017.201.

Evers-Vermeul, J., & Sanders, T. J. (2009). The emergence of Dutch connectives: How cumulative cognitive complexity explains the order of acquisition. *Journal of Child Language*, *36*(4), 829–854. doi:10.1017/S0305000908009227.

Ferreira, F., & Chantavarin, S. (2018). Integration and prediction in language processing: A synthesis of old and new. *Current Directions in Psychological Science*, *27*(6), 443–448. doi:10.1177/096372141879449.

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLOS One*, *8*(10), 1–18. doi:10.1371/journal.pone.0077661.

Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software*, *8*(15), 1–27. doi:10.18637/jss.v008.i15.

Frank, A. F., & Jaeger, T. F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (pp. 939–944). volume 30.

Freebody, P., & Anderson, R. C. (1983). Effects of vocabulary difficulty, text cohesion, and schema availability on reading comprehension. *Reading Research Quarterly*, *18*(3), 277–294. doi:10.2307/747389.

Frinsel, F. F., & Christiansen, M. H. (2024). Capturing individual differences in sentence processing: How reliable is the self-paced reading task? *Behavior Research Methods*, *56*, 6248–6257. doi:10.3758/s13428-024-02355-x.

Gerken, L., Wilson, R., & Lewis, W. (2005). Infants can use distributional cues to form syntactic categories. *Journal of Child Language*, *32*(2), 249–268. doi:10.1017/S0305000904006786.

Giulianelli, M., Wallbridge, S., & Fernández, R. (2023). Information value: Measuring utterance predictability as distance from plausible alternatives. In H. Bouamor, J. Pino, & K. Bali (Eds.), *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (pp. 5633–5653). Singapore: Association for Computational Linguistics. doi:10.18653/v1/2023.emnlp-main.343.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*(3), 371–395.

Grice, H. P. (1975). Logic and conversation. In P. Cole, & J. L. Morgan (Eds.), *Speech acts* (pp. 41–58). Leiden: Brill. doi:10.1163/9789004368811_003.

Grisot, C., & Blochowiak, J. (2017). Temporal connectives and verbal tenses as processing instructions: Evidence from French. *Pragmatics & Cognition*, *24*(3), 404–440. doi:10.1075/pc.17009.gri.

Grisot, C., & Blochowiak, J. (2021). Temporal relations at the sentence and text genre level: The role of linguistic cueing and non-linguistic biases—an annotation study of a bilingual corpus. *Corpus Pragmatics*, (pp. 1–41). doi:10.1007/s41701-021-00104-5.

Günther, F., Dudschig, C., & Kaup, B. (2016). Predicting lexical priming effects from distributional semantic similarities: A replication with extension. *Frontiers in Psychology*, *7*, 1646. doi:10.3389/fpsyg.2016.01646.

Gwet, K. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology*, *61*(1), 29–48. doi:10.1348/000711006X126600.

Haller, P., Koncic, I., Reich, D., & Jäger, L. A. (2023). Measurement reliability of individual differences in sentence processing: A cross-methodological reading corpus and Bayesian analysis.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Heilbron, M., Armeni, K., Schoffelen, J.-M., Hagoort, P., & De Lange, F. P. (2022). A hierarchy of linguistic predictions during natural language comprehension. *Proceedings of the National Academy of Sciences*, *119*(32), e2201968119. doi:10.1073/pnas.2201968119.

Hewett, F. (2023). APA-RST: A text simplification corpus with RST annotations. In M. Strube, C. Braud, C. Hardmeier, J. J. Li, S. Loaiciga, & A. Zeldes (Eds.), *Proceedings of the 4th Workshop on Computational Approaches to Discourse (CODI 2023)* (pp. 173–179). Toronto, Canada: Association for Computational Linguistics. doi:10.18653/v1/2023.codi-1.23.

Hinnell, J. (2019). The verbal-kinesic enactment of contrast in North American English. *The American Journal of Semiotics*, *35*(1-2), 55–92. doi:10.5840/ajs20198754.

Hobbs, J. R. (1979). Coherence and coreference. *Cognitive Science*, *3*(1), 67–90. doi:10.1207/s15516709cog0301$_4$.

Hoek, J., Evers-Vermeul, J., & Sanders, T. J. M. (2018). Segmenting discourse: Incorporating interpretation into segmentation? *Corpus Linguistics and Linguistic Theory*, *14*(2), 357–386. doi:10.1515/cllt-2016-0042.

Hoek, J., Evers-Vermeul, J., & Sanders, T. J. M. (2019a). Using the cognitive approach to coherence relations for discourse annotation. *Dialogue & Discourse*, *10*(2), 1–33. doi:10.5087/dad.2019.201.

Hoek, J., Rohde, H., Evers-Vermeul, J., & Sanders, T. J. M. (2021a). Expectations from relative clauses: Real-time coherence updates in discourse processing. *Cognition*, *210*, 104581. doi:10.1016/j.cognition.2020.104581.

Hoek, J., Rohde, H., Evers-Vermeul, J., & Sanders, T. J. M. (2021b). Scolding the child who threw the scissors: Shaping discourse expectations by restricting referents. *Language, Cognition and Neuroscience*, *36*(3), 382–399. doi:10.1080/23273798.2020.1852292.

Hoek, J., & Scholman, M. C. (2017). Evaluating discourse annotation: Some recent insights and new approaches. In *Proceedings of the 13th Joint ISO-ACL Workshop on Interoperable Semantic Annotation (ISA-13)* (pp. 1–13).

Hoek, J., & Scholman, M. C. (2023). Expressing non-volitional causality in english. In L. Jędrzejowski, & C. Fleczoreck (Eds.), *Micro- and Marcro-variation of Causal Clauses. Synchronic and Diachronic Insights* Studies in Language Companion Series (p. 167–183). Amsterdam: Benjamins.

Hoek, J., Scholman, M. C., & Sanders, T. J. (2021c). Is there less annotator agreement when the discourse relation is underspecified? *Integrating Perspectives on Discourse Annotation (DiscAnn)*, (pp. 1–6).

Hoek, J., Zufferey, S., Evers-Vermeul, J., & Sanders, T. J. M. (2017). Cognitive complexity and the linguistic marking of coherence relations: A parallel corpus study. *Journal of Pragmatics*, *121*, 113–131. doi:10.1016/j.pragma.2017.10.010.

Hoek, J., Zufferey, S., Evers-Vermeul, J., & Sanders, T. J. M. (2019b). The linguistic marking of coherence relations: Interactions between connectives and segment-internal elements. *Pragmatics & Cognition*, *25*(2), 276–309. doi:10.1075/pc.18016.hoe.

Hu, N., Chen, A., Li, F., Quené, H., & Sanders, T. J. M. (2022). A trade-off relationship between lexical and prosodic means in expressing subjective and objective causality: Evidence from English and Mandarin. In *Proceedings of Speech Prosody* (pp. 23–26).

Hu, N., Chen, A., Quené, H., & Sanders, T. J. M. (2023). The role of prosody in interpreting causality in English discourse. *PLOS One*, *18*(6), e0286003. doi:10.1371/journal.pone.0286003.

Huang, K.-J., Arehalli, S., Kugemoto, M., Muxica, C., Prasad, G., Dillon, B., & Linzen, T. (2024). Large-scale benchmark yields no evidence that language model surprisal explains syntactic disambiguation difficulty. *Journal of Memory and Language*, *137*, 104510. doi:10.1016/j.jml.2024.104510.

Jin, L., & de Marneffe, M.-C. (2015). The overall markedness of discourse relations. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (pp. 1114–1119).

Johnson, E. D., & Arnold, J. E. (2021). Individual differences in print exposure predict use of implicit causality in pronoun comprehension and referential prediction. *Frontiers in Psychology*, *12*, 2933. doi:10.3389/fpsyg.2021.672109.

Johnson, E. D., & Arnold, J. E. (2023). The frequency of referential patterns guides pronoun comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *49*(8), 1325. doi:10.1037/xlm0001137.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological review*, *87*(4), 329. doi:10.1037/0033-295X.87.4.329.

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

Keenan, J. M., Baillet, S. D., & Brown, P. (1984). The effects of causal cohesion on comprehension and memory. *Journal of Verbal Learning and Verbal Behavior*, *23*(2), 115–126. doi:10.1016/S0022-5371(84)90082-3.

Kehler, A. (2002). *Coherence, Reference, and the Theory of Grammar*. CSLI publications Stanford, CA.

Kehler, A. (2006). Discourse coherence. In L. Horn, & G. W. Ward (Eds.), *The handbook of pragmatics* (pp. 241–265). Oxford: Wiley Online Library.

Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics*, *25*(1), 1–44. doi:10.1093/jos/ffm018.

Kehler, A., & Rohde, H. (2017). Evaluating an expectation-driven question-under-discussion model of discourse interpretation. *Discourse Processes*, *54*(3), 219–238. doi:10.1080/0163853X.2016.1169069.

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

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363. doi:10.1037/0033-295X.85.5.363.

Kishimoto, Y., Sawada, S., Murawaki, Y., Kawahara, D., & Kurohashi, S. (2018). Improving crowdsourcing-based annotation of Japanese discourse relations. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

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. doi: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.

Knott, A., & Dale, R. (1994). Using linguistic phenomena to motivate a set of coherence relations. *Discourse Processes*, *18*(1), 35–62. doi:10.1080/01638539409544883.

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

Koornneef, A. (2021). The processing signature of anticipatory reading: An eye-tracking study on lexical predictions. *Linguistics*, *59*(2), 449–479. doi:10.1515/ling-2021-0014.

Koornneef, A. W., & Sanders, T. J. M. (2013). Establishing coherence relations in discourse: The influence of implicit causality and connectives on pronoun resolution. *Language and Cognitive Processes*, *28*(8), 1169–1206. doi:10.1080/01690965.2012.699076.

Kortmann, B. (2013). *Free adjuncts and absolutes in English: Problems of control and interpretation*. Routledge.

Kuperberg, G. R. (2016). Separate streams or probabilistic inference? What the N400 can tell us about the comprehension of events. *Language, Cognition and Neuroscience*, *31*(5), 602–616. doi:10.1080/23273798.2015.1130233.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. doi:10.1080/23273798.2015.1102299.

Kuperberg, G. R., Paczynski, M., & Ditman, T. (2011). Establishing causal coherence across sentences: An ERP study. *Journal of Cognitive Neuroscience*, *23*(5), 1230–1246. doi:10.1162/jocn.2010.21452.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161–163. doi:10.1038/307161a0.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. doi:10.18637/jss.v082.i13.

Laparle, S. M. (2022). *The shape of discourse: How gesture structures conversation*. Ph.D. thesis, University of California, Berkeley.

Lefcheck, J. S. (2016). piecewiseSEM: Piecewise structural equation modelling in r for ecology, evolution, and systematics. *Methods in Ecology and Evolution*, *7*(5), 573–579. doi:10.1111/2041-210X.12512.

Lenth, R. V. (2024). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.10.1.

Levy, R. P. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. doi:10.1016/j.cognition.2007.05.006.

Li, F., Mak, W. M., Evers-Vermeul, J., & Sanders, T. J. M. (2017). On the online effects of subjectivity encoded in causal connectives. *Review of Cognitive Linguistics*, *15*(1), 34–57. doi:10.1075/rcl.15.1.02li.

Linderholm, T., Everson, M. G., Van Den Broek, P., Mischinski, M., Crittenden, A., & Samuels, J. (2000). Effects of causal text revisions on more- and less-skilled readers' comprehension of easy and difficult texts. *Cognition and Instruction*, *18*(4), 525–556. doi:10.1207/S1532690XCI1804$_4$.

Lund, K., Burgess, C., & Atchley, R. A. (1995). Semantic and associative priming in high-dimensional semantic space. In *Proceedings of the 17th Annual conference of the Cognitive Science Society* (pp. 660–665). volume 17.

Mahowald, K., Ivanova, A. A., Blank, I. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko, E. (2024). Dissociating language and thought in large language models. *Trends in Cognitive Sciences*, *28*(6), 517–540. doi:10.1016/j.tics.2024.01.011.

Mak, W. M., & Sanders, T. J. (2013). The role of causality in discourse processing: Effects of expectation and coherence relations. *Language and Cognitive processes*, *28*(9), 1414–1437. doi:10.1080/01690965.2012.708423.

Von der Malsburg, T., & Angele, B. (2017). False positives and other statistical errors in standard analyses of eye movements in reading. *Journal of Memory and Language*, *94*, 119–133. doi:10.1016/j.jml.2016.10.003.

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, *92*, 57–78. doi:10.1016/j.jml.2016.04.001.

Mann, W. C., & Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse*, *8*(3), 243–281. doi:10.1515/text.1.1988.8.3.243.

Marchal, M., Scholman, M. C., & Demberg, V. (2021a). Clause type as a cue for processing discourse relations. doi:10.17605/OSF.IO/BKGF7.

Marchal, M., Scholman, M. C., & Demberg, V. (2021b). Semi-automatic discourse annotation in a low-resource language: Developing a connective lexicon for Nigerian Pidgin. In *Proceedings of the 2nd Workshop on Computational Approaches to Discourse* (pp. 84–94).

Marchal, M., Scholman, M. C., & Demberg, V. (2022a). Clause structure as a cue for discourse relations (coherence judgment). doi:10.17605/OSF.IO/CMF3S.

Marchal, M., Scholman, M. C., & Demberg, V. (2022b). Clause structure as a cue for on-line discourse processing. doi:10.17605/OSF.IO/U3GVS.

Marchal, M., Scholman, M. C., Yung, F., & Demberg, V. (2022c). Establishing annotation quality in multi-label annotations. In *Proceedings of the 29th International*

Conference on Computational Linguistics (pp. 3659–3668). Gyeongju, Republic of Korea: International Committee on Computational Linguistics.

Marx, E., Wittenberg, E., Marx, E., & Wittenberg, E. (2024). Eventuality type predicts temporal order inferences in discourse comprehension. *Glossa Psycholinguistics*, *3*(1), 1–24. doi:10.5070/G60116579.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. doi:10.1016/j.jml.2017.01.001.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101–B111. doi:10.1016/S0010-0277(01)00157-3.

McNamara, D. S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *55*(1), 51. doi:10.1037/h0087352.

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

McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*(3), 247–288. doi:10.1080/01638539609544975.

Mehti, R. (2024). *Beyond handcrafting: Can automatically generated screening questions replace handcrafted ones?*. Master's thesis, Saarland University.

Milburn, E., Warren, T., & Dickey, M. W. (2016). World knowledge affects prediction as quickly as selectional restrictions: Evidence from the visual world paradigm. *Language, Cognition and Neuroscience*, *31*(4), 536–548. doi:10.1080/23273798.2015.1117117.

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

Miltsakaki, E., Joshi, A., Prasad, R., & Webber, B. (2004). Annotating discourse connectives and their arguments. In *Proceedings of the Workshop Frontiers in Corpus Annotation at HLT-NAACL 2004* (pp. 9–16). Boston, MA, USA.

Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning*, *62*(1), 302–331. doi:10.1111/j.1467-9922.2010.00626.x.

Mulder, G. (2008). *Understanding causal coherence relations*. Ph.D. thesis, Utrecht University.

Münte, T. F., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature*, *395*(6697), 71–73. doi:10.1038/25731.

Murphy, M. V. (2022). *semEff: Automatic Calculation of Effects for Piecewise Structural Equation Models*. R package version 0.6.1, https://github.com/murphymv/semEff.

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

Myers, J. L., Shinjo, M., & Duffy, S. A. (1987). Degree of causal relatedness and memory. *Journal of Memory and Language*, *26*(4), 453. doi:10.1016/0749-596X(87)90101-X.

Nicenboim, B., Vasishth, S., Gattei, C., Sigman, M., & Kliegl, R. (2015). Working memory differences in long-distance dependency resolution. *Frontiers in Psychology*, *6*, 312. doi:10.3389/fpsyg.2015.00312.

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. doi:10.1044/jshr.3501.10.

Noordman, L. G., & Vonk, W. (1992). Readers' knowledge and the control of inferences in reading. *Language and Cognitive Processes*, *7*(3-4), 373–391. doi:10.1080/01690969208409392.

Noordman, L. G. M., & de Blijzer, F. (2000). On the processing of causal relations. In *Cause - Condition - Concession - Contrast: Cognitive and Discourse Perspectives* (pp. 35–56). Walter de Gruyter.

Noordman, L. G. M., & Vonk, W. (1998). Memory-based processing in understanding causal information. *Discourse Processes*, *26*(2-3), 191–212. doi:10.1080/01638539809545044.

Noordman, L. G. M., & Vonk, W. (2015). Inferences in discourse, psychology of. In *International Encyclopedia of the Social & Behavioral Sciences (2nd ed.) Vol. 12* (pp. 37–44). Elsevier.

Noordman, L. G. M., Vonk, W., Cozijn, R., & Frank, S. (2015). Causal inferences and world knowledge. In E. J. O'Brien, A. E. Cook, & R. F. Lorch Jr. (Eds.), *Inferences during reading* (p. 260). Cambridge: Cambridge University Press.

Noordman, L. G. M., Vonk, W., & Kempff, H. J. (1992). Causal inferences during the reading of expository texts. *Journal of Memory and Language*, *31*(5), 573–590. doi:10.1016/0749-596X(92)90029-W.

O'Reilly, T., & McNamara, D. S. (2007). Reversing the reverse cohesion effect: Good texts can be better for strategic, high-knowledge readers. *Discourse Processes*, *43*(2), 121–152. doi:10.1080/01638530709336895.

Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction*, *19*(3), 228–242. doi:10.1016/j.learninstruc.2008.04.003.

Paape, D., Vasishth, S., Paape, D., & Vasishth, S. (2022). Conscious rereading is confirmatory: Evidence from bidirectional self-paced reading. *Glossa Psycholinguistics*, *1*(1), 1–40. doi:10.5070/G6011182.

Pickering, M. J., & Majid, A. (2007). What are implicit causality and consequentiality? *Language and Cognitive Processes*, *22*(5), 780–788. doi:10.1080/01690960601119876.

Pinker, S., & Jackendoff, R. (2005). The faculty of language: What's special about it? *Cognition*, *95*(2), 201–236. doi:10.1016/j.cognition.2004.08.004.

Pitler, E., Louis, A., & Nenkova, A. (2009). Automatic sense prediction for implicit discourse relations in text. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP* (pp. 683–691).

Politzer-Ahles, S., Xiang, M., & Almeida, D. (2017). "before" and "after": Investigating the relationship between temporal connectives and chronological ordering using event-related potentials. *PLOS One*, *12*(4), e0175199. doi:10.1371/journal.pone.0175199.

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., & Webber, B. (2008). The Penn Discourse TreeBank 2.0. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*. Marrakech, Morocco: European Language Resources Association (ELRA).

Prasad, R., McRoy, S., Frid, N., Joshi, A., & Yu, H. (2011). The Biomedical Discourse Relation Bank. *BMC Bioinformatics*, *12*(1), 1–18. doi:10.1186/1471-2105-12-188.

Pusse, F., Sayeed, A., & Demberg, V. (2016). LingoTurk: Managing crowdsourced tasks for psycholinguistics. In *Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)* (pp. 57–61). San Diego, CA.

Pyatkin, V., Klein, A., Tsarfaty, R., & Dagan, I. (2020). QADiscourse: Discourse relations as qa pairs: Representation, crowdsourcing and baselines. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing* (pp. 2804–2819).

Pyatkin, V., Yung, F., Scholman, M. C., Tsarfaty, R., Dagan, I., & Demberg, V. (2023). Design choices for crowdsourcing implicit discourse relations: revealing the biases introduced by task design. *Transactions of the Association for Computational Linguistics*, *11*, 1014–1032. doi:10.1162/tacl$_{a0}$0586.

Quené, H., & Van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*(4), 413–425. doi:10.1016/j.jml.2008.02.002.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria.

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. doi:10.1037/0033-2909.124.3.372.

Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(2), 514.

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing* (pp. 3982–3992). Association for Computational Linguistics.

Robaldo, L., & Miltsakaki, E. (2014). Corpus-driven semantics of concession: Where do expectations come from? *Dialogue & Discourse*, *5*(1), 1–36. doi:10.5087/dad.2014.101.

Rohde, H., Dickinson, A., Schneider, N., Clark, C., Louis, A., & Webber, B. (2016). Filling in the blanks in understanding discourse adverbials: Consistency, conflict, and context-dependence in a crowdsourced elicitation task. In *Proceedings of the 10th Linguistic Annotation Workshop (LAW X)* (pp. 49–58). Berlin, Germany.

Rohde, H., Hoek, J., Keshev, M., & Franke, M. (2022). This better be interesting: A speaker's decision to speak cues listeners to expect informative content. *Open Mind*, *6*, 118–131. doi:10.1162/opmi$_{a0}$0058.

Rohde, H., & Horton, W. S. (2014). Anticipatory looks reveal expectations about discourse relations. *Cognition*, *133*(3), 667–691. doi:10.1016/j.cognition.2014.08.012.

Rohde, H., Tyler, J., & Carlson, K. (2017). Form and function: Optional complementizers reduce causal inferences. *Glossa (London)*, *2*(1). doi:10.5334/gjgl.134.

Rosa, E. C., & Arnold, J. E. (2017). Predictability affects production: Thematic roles can affect reference form selection. *Journal of Memory and Language*, *94*, 43–60. doi:10.1016/j.jml.2016.07.007.

RStudio Team (2020). *RStudio: Integrated Development Environment for R*. RStudio, PBC. Boston, MA.

Rysová, M., & Rysová, K. (2014). The centre and periphery of discourse connectives. In *Proceedings of the 28th Pacific Asia Conference on Language, Information and Computing* (pp. 452–459).

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. doi:10.1126/science.274.5294.1926.

Sanders, T. J. (1997). Semantic and pragmatic sources of coherence: On the categorization of coherence relations in context. *Discourse Processes*, *24*(1), 119–147. doi:10.1080/01638539709545009.

Sanders, T. J., & Evers-Vermeul, J. (2019). Subjectivity and causality in discourse and cognition. In *Empirical studies of the construction of discourse* (pp. 273–298). Amsterdam: John Benjamins Publishing Company.

Sanders, T. J., Land, J., & Mulder, G. (2007). Linguistics markers of coherence improve text comprehension in functional contexts. *Information Design Journal*, *15*(3), 219–235. doi:10.1075/idj.15.3.04san.

Sanders, T. J. M. (2005). Coherence, causality and cognitive complexity in discourse. In *Proceedings/Actes SEM-05, First International Symposium on the Exploration and Modelling of Meaning* (pp. 105–114). Toulouse, France.

Sanders, T. J. M., Demberg, V., Hoek, J., Scholman, M. C., Asr, F. T., Zufferey, S., & Evers-Vermeul, J. (2021). Unifying dimensions in coherence relations: How various annotation frameworks are related. *Corpus Linguistics and Linguistic Theory*, *17*(1), 1–71. doi:10.1515/cllt-2016-0078.

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. doi: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. doi:10.1080/01638539209544800.

Savic, O., Unger, L., & Sloutsky, V. M. (2022). Exposure to co-occurrence regularities in language drives semantic integration of new words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(7), 1064–1081. doi:10.1037/xlm0001122.

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human perception and performance*, *3*(1), 1. doi:10.1037/0096-1523.3.1.1.

Schler, J., Koppel, M., Argamon, S., & Pennebaker, J. W. (2006). Effects of age and gender on blogging. In *AAAI spring symposium: Computational approaches to analyzing weblogs* (pp. 199–205). volume 6.

Scholman, M., & Laparle, S. (accepted). Can gestures speak louder than words? The effect of gestural discourse markers on discourse expectations. *Discourse Processes*, (pp. 1–32).

Scholman, M. C. (2019). *Coherence relations in discourse and cognition: comparing approaches, annotations and interpretations*. Ph.D. thesis, Saarland University.

Scholman, M. C., Blything, L., Cain, K., Hoek, J., & Evers-Vermeul, J. (2022a). Discourse rules: The effects of clause order principles on the reading process. *Language, Cognition and Neuroscience*, *37*(10), 1277–1291. doi:10.1080/23273798.2022.2077971.

Scholman, M. C., & Demberg, V. (2017a). Crowdsourcing discourse interpretations: On the influence of context and the reliability of a connective insertion task. In *Proceedings of the 11th Linguistic Annotation Workshop (LAW)* (pp. 24–33). Valencia, Spain.

Scholman, M. C., & Demberg, V. (2017b). Examples and specifications that prove a point: Identifying elaborative and argumentative discourse relations. *Dialogue & Discourse*, *8*(2), 56–83. doi:10.5087/dad.2017.203.

Scholman, M. C., Demberg, V., & Sanders, T. J. M. (2020). Individual differences in expecting coherence relations: Exploring the variability in sensitivity to contextual signals in discourse. *Discourse Processes*, *57*(10), 844–861. doi:10.1080/0163853X.2020.1813492.

Scholman, M. C., Demberg, V., & Sanders, T. J. M. (2022b). Descriptively adequate and cognitively plausible? Validating distinctions between types of coherence relations. *Discours. Revue de linguistique, psycholinguistique et informatique. A journal of linguistics, psycholinguistics and computational linguistics*, *30*. doi:10.4000/discours.12075.

Scholman, M. C., Dong, T., Yung, F., & Demberg, V. (2022c). DiscoGeM: A crowdsourced corpus of genre-mixed implicit discourse relations. In *Proceedings of the Thirteenth International Conference on Language Resources and Evaluation (LREC'22)*. Marseille, France: European Language Resources Association (ELRA).

Scholman, M. C., Evers-Vermeul, J., & Sanders, T. J. (2016). Categories of coherence relations in discourse annotation: Towards a reliable categorization of coherence relations. *Dialogue & Discourse*, *7*(2), 1–28. doi:doi.org/10.5087/dad.2016.201.

Scholman, M. C., Marchal, M., Brown, A.-R., & Demberg, V. (submitted). DiscoNaija: A discourse-annotated parallel Nigerian Pidgin-English corpus.

Scholman, M. C., Marchal, M., & Demberg, V. (2024a). Connective comprehension in adults: The influence of lexical transparency, frequency, and individual differences. *Discourse Processes*, *61*(8), 381–403. doi:10.1080/0163853X.2024.2325262.

Scholman, M. C., Pyatkin, V., Yung, F., Dagan, I., Tsarfaty, R., & Demberg, V. (2022d). Design choices in crowdsourcing discourse relation annotations: The effect of worker selection and training. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference* (pp. 2148–2156).

Scholman, M. C., Rohde, H., & Demberg, V. (2017). "On the one hand" as a cue to anticipate upcoming discourse structure. *Journal of Memory and Language*, *97*, 47–60. doi:10.1016/j.jml.2017.07.010.

Scholman, M. C., Rohde, H., & Demberg, V. (2024b). Facilitation of lexical form or discourse relation: Evidence from contrastive pairs of discourse markers. *Glossa Psycholinguistics*, *3*(1). doi:10.5070/G60111353.

Schourup, L. (1999). Discourse markers. *Lingua*, *107*(3-4), 227–265. doi:10.1016/S0024-3841(96)90026-1.

Schütze, C. T., & Sprouse, J. (2013). Judgment data. In R. J. Podesva, & D. Sharma (Eds.), *Research Methods in Linguistics* (pp. 27–50). Cambridge: Cambridge University Press.

Schwab, J., & Liu, M. (2020). Lexical and contextual cue effects in discourse expectations: Experimenting with German 'zwar... aber'and English 'true/sure... but'. *Dialogue & Discourse*, *11*(2), 74–109. doi:10.5087/dad.2020.203.

Schwarz, F., & Zehr, J. (2021). Tutorial: Introduction to PCIbex–an open-science platform for online experiments: Design, data-collection and code-sharing. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (pp. 15–16). volume 43.

Segal, E. M., Duchan, J. F., & Scott, P. J. (1991). The role of interclausal connectives in narrative structuring: Evidence from adults' interpretations of simple stories. *Discourse Processes*, *14*(1), 27–54. doi:10.1080/01638539109544773.

Shipley, B. (2000). A new inferential test for path models based on directed acyclic graphs. *Structural Equation Modeling*, *7*(2), 206–218. doi:10.1207/S15328007SEM0702$_4$.

Sieker, J., Bott, O., Solstad, T., & Zarrieß, S. (2023). Beyond the bias: Unveiling the quality of implicit causality prompt continuations in language models. In *Proceedings of the 16th International Natural Language Generation Conference* (pp. 206–220).

Simner, J., & Pickering, M. J. (2005). Planning causes and consequences in discourse. *Journal of Memory and Language*, *52*(2), 226–239. doi:10.1016/j.jml.2004.04.006.

Škrjanec, I., Broy, F. Y., & Demberg, V. (2023). Expert-adapted language models improve the fit to reading times. In *Procedia Computer Science* (pp. 3488–3497). Elsevier volume 225.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302–319. doi:10.1016/j.cognition.2013.02.013.

Smith, R., Snow, P., Serry, T., & Hammond, L. (2021). The role of background knowledge in reading comprehension: A critical review. *Reading Psychology*, *42*(3), 214–240. doi:10.1080/02702711.2021.1888348.

Spooren, W. (1997). The processing of underspecified coherence relations. *Discourse Processes*, *24*(1), 149–168. doi:10.1080/01638539709545010.

Spooren, W., & Degand, L. (2010). Coding coherence relations: Reliability and validity. *Corpus Linguistics and Linguistic Theory*, *6*, 241–266.

Spooren, W., & Sanders, T. J. (2008). The acquisition order of coherence relations: On cognitive complexity in discourse. *Journal of pragmatics*, *40*(12), 2003–2026.

Sporleder, C., & Lascarides, A. (2008). Using automatically labelled examples to classify rhetorical relations: An assessment. *Natural Language Engineering*, *14*(3), 369–416. doi:10.1017/S1351324906004451.

Staub, A. (2015). The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. *Language and Linguistics Compass*, *9*(8), 311–327. doi:10.1111/lnc3.12151.

Staub, A. (2021). How reliable are individual differences in eye movements in reading? *Journal of Memory and Language*, *116*, 1–18. doi:10.1016/j.jml.2020.104190.

Staub, A., Grant, M., Astheimer, L., & Cohen, A. (2015). The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language*, *82*, 1–17. doi:10.1016/j.jml.2015.02.004.

Stede, M., Scheffler, T., & Mendes, A. (2018). Connective-Lex: A web-based multilingual lexical resource for connectives. *Discours. Revue de linguistique, psycholinguistique et informatique. A journal of linguistics, psycholinguistics and computational linguistics*, *24*. doi:10.4000/discours.10098.

Sweetser, E. (1990). *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure* volume 54. Cambridge University Press.

Taboada, M., & Mann, W. C. (2006). Rhetorical Structure Theory: Looking back and moving ahead. *Discourse Studies*, *8*(3), 423–459. doi:10.1177/1461445606061881.

The Editors of Discourse Processes (2024). Retraction statement - HDSP/ Blumenthal-Dramé Article. doi:10.1080/0163853X.2024.2348297.

Trabasso, T., & Van Den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, *24*(5), 612–630. doi:10.1016/0749-596X(85)90049-X.

Traxler, M. J., Sanford, A. J., Aked, J. P., & Moxey, L. M. (1997). Processing causal and diagnostic statements in discourse. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(1), 88.

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

Tskhovrebova, E., Zufferey, S., & Gygax, P. (2023). Exploring the sensitivity to alternative signals of coherence relations: The case of French speaking teenagers. *Dialogue & Discourse*, *14*(2), 49–82. doi:10.5210/dad.2023.202.

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

Tyler, J. (2013). Prosodic correlates of discourse boundaries and hierarchy in discourse production. *Lingua*, *133*, 101–126. doi:10.1016/j.lingua.2013.04.005.

Unger, L., Vales, C., & Fisher, A. V. (2020). The role of co-occurrence statistics in developing semantic knowledge. *Cognitive Science*, *44*(9), e12894. doi:10.1111/cogs.12894.

Van Enschot, R., Spooren, W., van den Bosch, A., Burgers, C., Degand, L., Evers-Vermeul, J., Kunneman, F., Liebrecht, C., Linders, Y., Maes, A. et al. (2024). Taming our wild data: On intercoder reliability in discourse research. *Dutch Journal of Applied Linguistics*, *13*. doi:10.51751/dujal16248.

Van Silfhout, G., Evers-Vermeul, J., Mak, W. M., & Sanders, T. J. M. (2014). Connectives and layout as processing signals: How textual features affect students' processing and text representation. *Journal of Educational Psychology*, *106*(4), 1036. doi:10.1080/0163853X.2014.905237.

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. doi:10.1080/0163853X.2014.905237.

Vasishth, S., von der Malsburg, T., & Engelmann, F. (2013). What eye movements can tell us about sentence comprehension. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(2), 125–134. doi:10.1002/wcs.1209.

Venhuizen, N. J., Crocker, M. W., & Brouwer, H. (2019). Expectation-based comprehension: Modeling the interaction of world knowledge and linguistic experience. *Discourse Processes*, *56*(3), 229–255. doi:10.1080/0163853X.2018.1448677.

Wan, S., Bourgonje, P., Xiao, H., & Ho, C. W. C. (2024). Chinese-DiMLex: a lexicon of Chinese discourse connectives. *Language Resources and Evaluation*, *59*, 1–22. doi:10.1007/s10579-024-09761-9.

Webber, B. (2013). What excludes an alternative in coherence relations? In *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013)–Long Papers* (pp. 276–287).

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

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

Wetzel, M., Tskhovrebova, E., Gygax, P. M., & Zufferey, S. (2023). Pragmatic and syntactic constraints on French causal connectives: An evaluation of native and non-native speakers' sensitivity. *Journal of Pragmatics*, *209*, 89–107. doi:10.1016/j.pragma.2023.03.001.

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. doi:10.3390/languages5030035.

Wetzel, M., Zufferey, S., & Gygax, P. M. (2021). How robust is discourse processing for native readers? the role of connectives and the coherence relations they convey. *Frontiers in Psychology*, *13*, 459. doi:10.3389/fpsyg.2022.822151.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wilcox, E. G., Ding, C., Sachan, M., & Jäger, L. A. (2024). Mouse tracking for reading (motr): A new naturalistic incremental processing measurement tool. *Journal of Memory and Language*, *138*, 104534. doi:10.1016/j.jml.2024.104534.

Wilcox, E. G., Pimentel, T., Meister, C., Cotterell, R., & Levy, R. P. (2023). Testing the predictions of surprisal theory in 11 languages. *Transactions of the Association for Computational Linguistics*, *11*, 1451–1470. doi:10.1162/tacl$_{a0}$0612.

Wong, R., Reichle, E. D., & Veldre, A. (2024). Prediction in reading: A review of predictability effects, their theoretical implications, and beyond. *Psychonomic Bulletin & Review*, . doi:10.3758/s13423-024-02588-z.

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

Xiao, H., van Hout, R. W. N. M., Sanders, T. J. M., & Spooren, W. P. M. S. (2021). A cognitive account of subjectivity put to the test: Using an insertion task to investigate Mandarin result connectives. *Cognitive Linguistics*, *32*(4), 671–702. doi:10.1515/cog-2020-0075.

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. doi:10.1080/0163853X.2016.1272088.

Xu, X., Jiang, X., & Zhou, X. (2015). When a causal assumption is not satisfied by reality: Differential brain responses to concessive and causal relations during sentence comprehension. *Language, Cognition and Neuroscience*, *30*(6), 704–715. doi:10.1080/23273798.2015.1005636.

Yao, R., Husband, E. M., & Altshuler, D. (2024). Implicit questions-under-discussion raise expectations only in at-issue main clause. 37th Annual Conference on Human Sentence Processing.

Yi, E., & Koenig, J.-P. (2021). Grammar modulates discourse expectations: Evidence from causal relations in English and Korean. *Language and Cognition*, *13*(1), 99–127. doi:10.1017/langcog.2020.29.

Yung, F., Ahmad, M., Scholman, M. C., & Demberg, V. (2024a). Prompting implicit discourse relation annotation. In S. Henning, & M. Stede (Eds.), *Proceedings of The 18th Linguistic Annotation Workshop (LAW-XVIII)* (pp. 150–165). St. Julians, Malta: Association for Computational Linguistics.

Yung, F., Anuranjana, K., Scholman, M. C., & Demberg, V. (2022). Label distributions help implicit discourse relation classification. In *Proceedings of the 3rd Workshop on Computational Approaches to Discourse* (pp. 48–53). Gyeongju, Republic of Korea and Online: International Conference on Computational Linguistics.

Yung, F., Demberg, V., & Scholman, M. C. (2019). Crowdsourcing discourse relation annotations by a two-step connective insertion task. In *Proceedings of the 13th Linguistic Annotation Workshop (LAW)* (pp. 16–25).

Yung, F., Duh, K., Komura, T., & Matsumoto, Y. (2017). A psycholinguistic model for the marking of discourse relations. *Dialogue and Discourse*, *8*(1), 106–131. doi:10.5087/dad.2017.104.

Yung, F., Scholman, M. C., Lapshinova-Koltunski, E., Pollkläsener, C., & Demberg, V. (2023). Investigating explicitation of discourse connectives in translation using automatic annotations. In S. Stoyanchev, S. Joty, D. Schlangen, O. Dusek, C. Kennington, & M. Alikhani (Eds.), *Proceedings of the 24th Annual Meeting of the Special Interest Group on Discourse and Dialogue* (pp. 21–30). Prague, Czechia: Association for Computational Linguistics. doi:10.18653/v1/2023.sigdial-1.2.

Yung, F., Scholman, M. C., Zikánová, Š., & Demberg, V. (2024b). DiscoGeM 2.0: A Parallel Corpus of English, German, French and Czech Implicit Discourse Relations. In *Proceedings of the 2024 Joint International Conference on Computational*

Linguistics, Language Resources and Evaluation (LREC-COLING 2024) (pp. 4940–4956).

Zeyrek, D., Mendes, A., Grishina, Y., Kurfalı, M., Gibbon, S., & Ogrodniczuk, M. (2020). TED Multilingual Discourse Bank (TED-MDB): A parallel corpus annotated in the PDTB style. *Language Resources and Evaluation*, *54*(2), 587–613. doi:10.1007/s10579-019-09445-9.

Zufferey, S. (2016). Discourse connectives across languages: Factors influencing their explicit or implicit translation. *Languages in Contrast*, *16*(2), 264–279. doi:10.1075/lic.16.2.05zuf.

Zufferey, S., & Degand, L. (2017). Annotating the meaning of discourse connectives in multilingual corpora. *Corpus Linguistics and Linguistic Theory*, *13*(2), 399–422. doi:10.1515/cllt-2013-0022.

Zufferey, S., & Degand, L. (2024). *Connectives and Discourse Relations*. Cambridge: Cambridge University Press.

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

Zufferey, S., & Gygax, P. M. (2016). The role of perspective shifts for processing and translating discourse relations. *Discourse Processes*, *53*(7), 532–555. doi:10.1080/0163853X.2015.1062839.

Zwaan, R. A., & Rapp, D. N. (2006). Discourse comprehension. In M. J. Traxler, & M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics* (pp. 725–764). Elsevier. doi:10.1016/B978-012369374-7/50019-5.