Sentence Processing at the Interfaces

–

Dependency Relations Across Levels of Representation

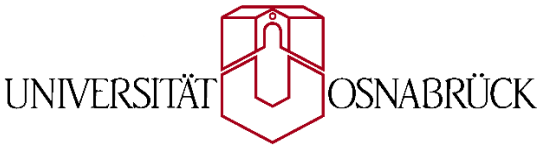
Dissertation

zur Erlangung des Grades

Doktor der Philosophie (Dr. phil.)

im Fachbereich Humanwissenschaften der

Universität Osnabrück



vorgelegt von

Juliane Schwab

Osnabrück, Juni 2022

SCIENTIFIC SUPERVISORS:
Prof. Dr. Mingya Liu
Prof. Dr. Jutta L. Mueller

NON-SCIENTIFIC SUPERVISOR:
Dr. phil. habil. Annette Hohenberger

CO-EXAMINERS:
Prof. Dr. Ming Xiang
Prof. Dr. Peter König

ABSTRACT (English)

One of the hallmark features of language is that linguistic elements (from morphemes to words and phrases) commonly relate to each other, thus instantiating a dependency relation between different parts of the linguistic input. Within psycholinguistics, key questions have been how comprehenders keep track of such dependencies and what (top-down and bottom-up) cognitive mechanisms are engaged during the resolution of dependency relations. In this thesis, I investigate the processing and comprehension of three different types of dependency relations for a state-of-the-art perspective onto the shared, and the unique, mechanisms engaged to resolve linguistic dependency relations. The thesis has two major components: In its first part, focusing on expectation-based processing mechanisms, it provides novel empirical evidence for the engagement of expectations in adults' processing of dependencies at various levels of linguistic representation—from the syntactic dependency between German determiners and relative clauses to the discourse-level dependency relation in concessive discourse relations. In its second part, focused on the particular dependency relation between polarity sensitive expressions and their (anti-)licenser, the thesis combines theoretical and empirical linguistic perspectives to illuminate the mechanisms involved in the licensing, comprehension, processing, and acquisition of polarity sensitive expressions. Therein, the included studies demonstrate comprehenders' immediate sensitivity to semantic and pragmatic properties of the sentential contexts in which negative polarity items appear, and attest to dependency-specific memory effects in adults' processing and comprehension. From the developmental perspective, these findings are complemented by a study showing that the acquisition of polarity sensitive expressions extends at least into early adolescence, with substantial variation that may be due to pragmatic development and input-related differences. The combined findings from both sections of this thesis have impact on linguistic theory and psycholinguistics alike.

ABSTRACT (German)

Eines der Hauptcharakterisitika von Sprache ist, dass sprachliche Elemente (von Morphemen über Wörter bis hin zu Phrasen) in der Regel miteinander in Beziehung stehen, sodass eine Abhängigkeitsbeziehung zwischen verschiedenen Teilen des sprachlichen Inputs besteht. Schlüsselfragen in der Psycholinguistik sind, wie solche Abhängigkeiten während des Sprachverständnisses im Auge behalten werden und welche kognitiven Mechanismen („top-down“ und „bottom-up“) zur Auflösung von Abhängigkeitsbeziehungen zum Einsatz kommen. In der vorliegenden Arbeit untersuche ich die Verarbeitung drei verschiedener Arten von Abhängigkeitsbeziehungen, für einen aktuellen Überblick über die gemeinsamen und die einzigartigen Mechanismen, die bei der Auflösung sprachlicher Abhängigkeiten zum Einsatz kommen. Die Arbeit besteht aus zwei Hauptkomponenten: In ihrem ersten Teil werden neue empirische Belege dafür präsentiert, dass Erwachsene zur Verarbeitung von Abhängigkeiten auf verschiedenen Ebenen sprachlicher Repräsentation erwartungsbasierte Mechanismen verwenden—von der syntaktischen Abhängigkeitsbeziehung zwischen deutschen Determinierern und Relativsätzen bis hin zu Diskursabhängigkeiten in konzessiven Diskursbeziehungen. In ihrem zweiten Teil, der sich auf die besondere Abhängigkeitsbeziehung zwischen polaritätssensitiven Ausdrücken und ihren (Anti-)Lizensoren konzentriert, kombiniert die Arbeit theoretische und empirisch-linguistische Perspektiven, um die Mechanismen zu beleuchten, die an der Lizensierung, dem Verständnis, der Verarbeitung, und der Akquisition polaritätssensitiver Ausdrücke beteiligt sind. Darin demonstrieren die beinhalteten Studien eine unmittelbare Sensibilität während des Sprachverständnisses für semantische und pragmatische Eigenschaften der Satzkontexte, in denen negative Polaritätsausdrücke vorkommen, und belegen zudem abhängigkeitsspezifische Gedächtniseffekte bei der Verarbeitung und dem Verständnis von Erwachsenen. Aus entwicklungspsychologischer Sicht werden diese Befunde durch eine Studie ergänzt, die zeigt, dass sich der Erwerb polaritätssensitiver Ausdrücke mindestens bis ins frühe Jugendalter erstreckt, wobei erhebliche Unterschiede aufgezeigt werden, die möglicherweise auf die pragmatische Entwicklung und inputbezogene Unterschiede zurückzuführen sind. Die kombinierten Ergebnisse aus beiden Abschnitten dieser Arbeit haben Implikationen für die Sprachtheorie und die Psycholinguistik gleichermaßen.

ACKNOWLEGMENTS

When I started studying cognitive science, I admittedly did not go into with any particular interest in language. My fascination with linguistics came into being only through the genuine enthusiasm of my teachers and supervisors, Mingya Liu and Jutta Mueller, as well as that of my first linguistics professor and former boss, Peter Bosch. I am deeply grateful for all that they have taught me, and for the continuous mentoring, encouragement, and support Mingya and Jutta have offered. I have come to know you both as excellent, rigorous scientists, but also as exceedingly kind and caring people. I hope that we will stay in touch and that we will continue working together for many years to come.

I would also like to thank my co-author Ming Xiang for being extraordinarily supportive as mentor and advisor, and for lending her time, expertise, and trust in our research project.

During the last few years, I have greatly benefitted from scientific and non-scientific exchanges with my fellow doctoral students in Osnabrück, Berlin, and Vienna, most of all my former office mate Ivonne Weyers, as well as Priscila Borba Borges, Britta Grusdt, Xenia Ohmer, Georg Schroeter, and Vinicius Macuch Silva. Further, I am thankful for insightful discussions with Caroline Beese, Danielle Benesch, Rachel Bouserhal, Carolin Dudschig, Thomas Gruber, Annette Hohenberger, and Barbara Kaup.

Beyond the world of science, I am thankful to my family and my friends, whom I'll spare the honour of being named, for their encouragement, their open minds and hearts, and for distracting me (rightfully so!) from work once in a while.

Finally, Mohamed, thank you for everything—that which I *can* put into words and that which I often *can't*. I know I have put you through a lot this year. Thank you for showing me patience and kindness, and for expanding my world.

I dedicate this thesis to my mother, for all that she is in my life and all the love she holds.

"Zu fragmentarisch ist Welt und Leben!
Ich will mich zum deutschen Professor begeben,
Der weiß das Leben zusammenzusetzen,
Und er macht ein verständliches System daraus;
Mit seinen Nachtmützen und Schlafrockfetzen
Stopft er die Lücken des Weltenbaus."

— Heinrich Heine

CONTENTS

PART I

Introduction and Background

"We demand rigidly defined areas of doubt and uncertainty!"

— Vroomfondel, in Douglas Adams' "The Hitchhiker's Guide to the Galaxy"

I.1: INTRODUCTION

It is considered one of the hallmark features of natural language that linguistic elements at various levels of representation form dependency relations with one another. This is to say that for two (or more) dependent elements *a* and *b* within a continuous linguistic input stream as in (1), occurrence of *a* predicts the later occurrence of *b*, or occurrence of *b* necessitates the prior occurrence of *a*, sometimes both. If only one of the two elements requires the respective other one, the dependency relation is asymmetric; if both require each other, the dependency relation is symmetric. Dependency relations exist at various levels of representation, from phonemes, to morphology, syntax, semantics, and even discourse. The focus within this thesis will be on dependency relations in syntax, discourse, and the interface of syntax, semantics, and pragmatics.

(1) **Asymmetric dependency relations**

x x ***a*** *x x x b x x*

a needs *b*

x x a x x x ***b*** *x x*

b needs *a*

Symmetric dependency relations

x x ***a*** *x x x* ***b*** *x x*

a needs *b*

b needs *a*

Symmetric (agreement) dependencies are attested, among others, in morphological and syntactic relations; to formally mark grammatical relations on the noun (Blake, 2001), for instance, languages like German use case inflections on the determiner and the noun (in addition to number and gender features). Both noun phrase (NP) constituents must carry appropriate matching case marking for the phrase to be well-formed (2). Moreover, at the syntactic level, sentential subjects within a sentence are dependent on their verbal head, which in turn (minimally, assuming the verb is intransitive,) requires a subject argument. In languages that employ such marking on the verb, subject and verb have to agree in their relevant morphosyntactic features, e.g., number in (2).

(2) *Das Leben eines Hasens ist kurz.*
The$_{\text{NOM.SING.NEUT}}$ life$_{\text{NOM.SING.NEUT}}$ a$_{\text{GEN.SING.MASC}}$ hare$_{\text{GEN.SING.MASC}}$ is$_{\text{SING}}$ short
'A hare's life is short.'

Asymmetric dependency relations, conversely, are abound in a variety of semantic, syntactic, pragmatic, and discourse phenomena. Staying on the issue of subject-verb dependencies, verbs like *streunern* 'stray' in (3), for instance, require an animate

subject. Animacy, as a semantic feature of nouns, thus constrains whether a noun can function as argument of the verb. However, contrary to the agreement dependencies discussed above, animacy is not a feature marked on the verb itself. Syntactically, asymmetric dependency relations exist, for instance, between the referent of noun phrases such as *die Katze* 'the cat' in (3) that are (optionally) restricted by a relative clause (RC) modifier. This RC is dependent on the noun head it modifies; on the other hand, bare noun phrases themselves do not require a relative clause and indeed rarely take one (see **Chapter II.1**). The relative clause is thus asymmetrically dependent on its head noun. At the discourse level, example (3) additionally demonstrates that the two propositions within the sentence relate to each other via a concessive discourse relation, that is, the sentence involves the assertion of two propositions *p* and *q* which are presupposed to be incompatible with one another (Konig & Siemund, 2000). In the provided example, this discourse relation is indicated both by the connective *aber* 'but' and by the additional discourse marker *zwar* 'true' on the conceded argument. The latter is completely optional, such that the establishment of a concessive discourse relation is not dependent on marking with *zwar*. It is part of the core lexical meaning of *zwar*, however, that it functions as marker of concessivity (see **Chapter II.2**). Therefore, occurrence of *zwar* on *p* asymmetrically requires the subsequent appearance of a proposition *q* functioning as second argument of a concessive discourse relation.

(3) *Die Katze(, die durch die Nachbarschaft streunert,) ist (zwar) wild, aber auch sehr freundlich.*
The cat that through the neighbourhood strays is true wild but also very friendly
'(True,)the cat (that strays through the neighbourhood) is feral, but it's also very friendly.'

Owing back to Chomsky's development of transformational generative grammar (Chomsky, 1956, 1957, 1965), the formation of structural dependency relations has been assumed to be a defining feature of natural language, allowing for the computation of an infinite number of novel utterances from a limited lexical inventory and a small set of combinatorial rules. This feature, also termed the recursiveness of natural language, has sometimes been argued to be the crucial factor separating natural language from all other forms of (animal) communication systems (Chomsky, 2010; Hauser et al., 2002).

In light of this background, it is not surprising that a considerable amount of the psycholinguistic literature is concerned with dependency phenomena. Although they are not the focus of this thesis, symmetric dependency relations like subject-verb agreement, for instance, have been discussed with respect to the representation of morphological features on (complex) noun phrases (Bock et al., 2001; Eberhard et al., 2005; Vigliocco et al., 1996), the computational mechanisms that match the agreement features of noun and verb (Badecker & Kuminiak, 2007; Jäger et al., 2020), and the factors underlying comprehenders' and producers' susceptibility to agreement errors like in '*The key to the cabinets is/*are rusty'* (e.g., Clifton Jr. et al., 1999; Dillon et al.,

2013; Franck et al., 2002; Hammerly et al., 2019; Jäger et al., 2020; Parker & Phillips, 2016; Patson & Husband, 2016; Pearlmutter et al., 1999; Staub, 2009; Vigliocco & Nicol, 1998). By contrast, an equally large body of work on asymmetric dependency relations, primarily focusing on syntactic dependency relations, has highlighted the cognitive computational mechanisms engaged during the establishment of such dependency relations at a sentence level (Engelmann et al., 2019; Futrell et al., 2020; Gibson, 2000; R. Levy, 2008; Lewis & Vasishth, 2005; see below).

As outlined at the outset of this section, asymmetric dependency relations comprise retrospective ones, where the dependent element needs to establish a relation with an element that has previously been encountered (the *b*-element needs the *a*-element), and prospective ones, i.e., those where the dependent element needs to establish a relation with an element that has yet to appear in the linguistic input (the *a*-element predicts the *b*-element). Crucially, these two types of dependency relations differ in the demands they place on the cognitive system: for the first, encountering the *b*-element in the linguistic input triggers a memory search for the *a*-element that is required to license the appearance of *b* (e.g., Lewis & Vasishth, 2005). At a minimum, this requires retrieval of stored representations at the level of granularity appropriate for the respective dependency relation (e.g., words in the local sentential context or parts of the preceding discourse segment). For the second, encountering the *a*-element does not yet guarantee that the required *b*-element will appear. Therefore, it has been argued that prospective dependency relations involve an expectation (also conceptualised as prediction or as "active dependency formation", particularly in the case of filler-gap dependencies, Aoshima et al., 2002; Frazier & Flores d'Arcais, 1989; Omaki et al., 2015; Stowe, 1986) for the pending element (Levy, 2008). Although the engagement of these memory- and expectation-based processes are not exclusive to asymmetric dependency relations, asymmetric dependencies highlight the respective contribution of one mechanism over the other, thus rendering them a prime candidate to study memory and expectations in language processing.

An assumption often implicit in the literature on the aforementioned processing mechanisms is that they constitute general cognitive processing constraints,[1] and therefore target linguistic representations at all relevant processing levels. At the same time, however, dependency relations at various levels of representation differ in critical linguistic properties that may affect their processing—that is, they might differ in ways that affect the manner and extent to which the involved representations are subject to memory decay and interference, and to which the relevant representations are employed in the generation of expectations for upcoming elements. Neurophysiological studies on linguistic prediction, for instance, have suggested that word-level event-related potential (ERP) responses to confirmed and violated predictions differ in their

[1] Another cognitive constraint on processing (language, and perceptual input in general) is imposed by limited attentional resources. The role of attention in sentence processing has been discussed with respect to different processing styles, such that comprehenders may not attend to all parts of the language input equally. Processing under minimised attention to the language input has been termed superficial or "good-enough" processing (Ferreira et al., 2002; Ferreira & Lowder, 2016).

temporospatial signature depending on the level of representation the prediction was formed at (Bovolenta & Husband, 2021; Kuperberg et al., 2019). Kuperberg et al. (2019) find that semantic prediction violations result in an increased N400 (see **Chapter I.2**), whereas lexical (event) prediction violations elicit a late frontal positivity, and event structure (animacy) prediction violations result in a late posterior positivity or P600. Bovolenta and Husband (2021) additionally find evidence for structural prediction effects, resulting in a frontal negativity for violations of phrase structure predictions. The precise functional roles of these electrophysiological signatures are debated within the field; in any case, however, the divergent patterns observed here suggest that the neural mechanisms involved in prediction are sensitive to differences in the linguistic level of representation.

In addition, with respect to the processing mechanisms taking place between two dependent elements, it has been suggested that particularly strong lexical predictions for upcoming words (as in (4)) may result in what is termed "pre-updating" the sentence representation (as compared to merely "pre-activating" the representation of the predicted word) (Kuperberg & Jaeger, 2016; Ness & Meltzer-Asscher, 2021). Ness and Meltzer-Asscher (2021) conceptualise this as a "race" process wherein words that are pre-activated probabilistically based on the sentence context (but also lexical frequency and random noise) compete for sufficient activation to be retrieved. Once a word reaches a threshold activation level, it is integrated into the sentence representation that is maintained in working memory (pre-updating). The extent to which such mechanisms could generalise to expectations (or predictions) that concern higher levels of representation—at which expectations do not concern lexical items, but pragmatic properties of the sentence or the expected sentence or discourse structure (see **Chapter I.3**)—is unclear.

(4) The day was breezy so the boy went outside to fly a… {$\text{kite}_{\text{predicted}}$/$\text{plane}_{\text{unpredicted}}$}

Finally, memory constraints, too, may exert unique contributions depending on the level of representation at which a dependency relation operates. For instance, there is ample evidence in syntactic processing for memory interference from stored representations that share semantic or syntactic features with the target for memory retrieval (Clifton Jr. et al., 1999; Dillon et al., 2013; Franck et al., 2002; Hammerly et al., 2019; Jäger et al., 2020; Martin & McElree, 2009; Parker, 2018; Patson & Husband, 2016; Pearlmutter et al., 1999; Staub, 2009b; Tan et al., 2017; Van Dyke, 2007; Van Dyke & McElree, 2006; Vigliocco & Nicol, 1998), yet memory interference effects do not appear to generalise straightforwardly across linguistic dependencies. Mixed results for anaphora resolution (Dillon et al., 2013; Sturt, 2003; Xiang et al., 2009; but see Patil et al., 2016) and NPI licensing (Muller & Phillips, 2020; Orth et al., 2020; Parker & Phillips, 2016; **Chapter III.3**), for instance, suggest that memory interference effects depend on the type of grammatical dependency relation and may be more selective than predicted by current memory-based processing accounts. In addition, as shown in **Chapter III.4**, some dependency relations, such as that between a polarity sensitive expression and its licensing context, may be subject to idiosyncratic memory

interference effect from (polarity insensitive) lexical competitors, which do not instantiate such a dependency relation.

This thesis comprises two main parts addressing various aspects in the processing and comprehension of dependency relations at three levels of representation (see an overview in **Table I.1**), addressing both how expectations and memory affect their on-line processing and how their comprehension (both in adults and children) is affected by dependency-specific linguistic properties. In the first part (**Chapters II.1** and **II.2**), I investigate expectation-based mechanisms in the processing of dependency relations at the level of *syntax* and *discourse*. Therein, I provide evidence for (a) an expectation-based anti-locality effect (i.e., processing facilitation under increased dependency length) in the parsing of the dependency relation between a German determiner and RCs (as discussed for (3)) and (b) for independent effects of narrow lexical and broad pragmatic cues on discourse expectations and perceived discourse coherence in concessive relations, again illustrated in (3). These studies contribute empirical evidence that furthers our understanding of the levels of representation expectations are generated from, and of their downstream effects on processing, independent of potentially co-active memory-based mechanisms.

In the second part (**Chapters III.1–III.5**), I focus on the dependency relation between polarity sensitive expressions and their licensing context. This dependency relation is situated at the *interface* of *syntax, semantics,* and *pragmatics* (among others: Chierchia, 2004, 2006; Giannakidou, 1998, 2006; Israel, 1996, 2011; Kadmon & Landman, 1993; Krifka, 1995; Ladusaw, 1979). Its unique linguistic properties (see **Chapters I.3** and **III.1**) have proven challenging for formal linguistic descriptions of said relation (*ibid*), for psycholinguistic research concerned with on-line processing (among others: Liu et al., 2019; Orth et al., 2020; Parker & Phillips, 2016; Saddy et al., 2004; Szabolcsi et al., 2008; Vasishth et al., 2008; Xiang et al., 2013, 2016; Yurchenko et al., 2013) and postlicensing[2] effects (Liu, 2019), and for developmental research concerned with the mechanisms enabling children to acquire polarity sensitive expressions (Koster & van der Wal, 1995; Lin et al., 2015, 2018; Tieu & Lidz, 2016). To contribute to the field, I provide a broad-scale evaluation of NPI licensing, processing, comprehension, and acquisition, based firmly in a theoretical framework on the linguistic properties that define the peculiar dependency relation between NPI and licensing context. In doing so, I first argue for a scalar pragmatic account of NPI licensing (**Chapter III.1**), particularly with respect to so-called attenuating NPIs like English *all that* (5) or German *sonderlich* ('particularly'). The proposed analysis is subsequently supported by empirical studies showing processing interactions between the licensing requirement of attenuating NPIs and semantic and pragmatic properties of the sentential context, both with respect to intrusion from close-by scalar environments (**Chapter III.3**) and with respect to interactions with pragmatic properties of various types of conditionals (5b) (**Chapter III.2**). Further, I show that comprehension and

[2] Postlicensing effects refer to "interpretive effects by licensed NPIs in a sentence" (Liu, 2019: 4), e.g., whether NPIs introduce speaker bias.

acquisition of polarity sensitive expressions are subject to idiosyncratic influences, e.g., of item-specific input frequencies, lexical competition in memory, and individuals' language aptitude (**Chapters III.4** and **III.5**).

(5) a. Mary is*(n't) all that happy.

b. If Mary {were/#is} all that happy, she {would have stayed/#will stay} in her job.

Overall, I demonstrate that we stand to gain broader insight onto the cognitive mechanisms involved in language processing (as well as comprehension and acquisition) by widening our perspective on the empirical linguistic phenomena we investigate. My thesis reveals that all of the investigated dependency relations are subject to expectation and memory effects, but also display dependency-specific differences that prove informative about limitations of contemporary cognitive language processing models. As I show for NPIs, theoretical linguistics crucially provides the tools to adequately capture formal properties that define the studied dependency relations, which in turn act as explanatory variables about effects in language processing. In blending theoretical and psycholinguistic perspectives, employing cutting-edge theories and methods from both fields, my collective contributions have implications for linguistic theory and psycholinguistics alike.

Level of representation	Phenomenon	Chapter	(Psycho-)linguistic domain	Cognitive domain	Example sentences
Syntax	Determiner-RC relation	II.1	Processing Comprehension	Expectation Memory	*{**der/ derjenige**} tapfere Junge (…) , dessen Bein…* {**the/ the-one**} brave boy (…) , whose leg…
Discourse	Concessive discourse relations	II.2	Processing Comprehension	Expectation	*…joggt gerne (**draußen**). Er hat (**zwar**) ein Laufband…, aber er…* …jogs like (**outdoors**). He has (**true**) a treadmill…, but he..
Syntax-semantics-pragmatics interface	Polarity sensitive expressions	III.1	Theory Comprehension	Scalarity	*Wenn die Schüler **sonderlich** aufmerksam {#sind, werden…/ wären, würden…}* If the students **particularly**$_{NPI}$ attentive {be$_{IND}$, will…/ be$_{SBJV}$, would…}
		III.2	Processing Comprehension	Pragmatic inference Scalarity	*1. Hypothetical indicative/counterfactual conditional:* If the students {#have/had} been **all that** attentive during class, … *2. Hypothetical indicative/premise conditional:* {A: The students have been very attentive... / #A: The students will start their exam season…} B: If the students have been **all that** attentive, …
		III.3	Processing	Memory Scalarity	*{*Der/ Kein} Bauer, der {das/ kein} Pestizid verwendete,* {*The/ No} farmer, who {the/ no} pesticide used, *war **{jemals/ so recht}**…* was {**ever**$_{NPI}$/ **really**$_{NPI}$}…
		III.4	Processing Comprehension	Memory Individual differences	*{**Sonderlich/ So recht**} … {kein/ nicht/ *affirmative}…* {**Particularly**$_{NPI}$/ **Really**$_{NPI}$} … {no/ not/ *affirmative}…
		III.5	Acquisition Comprehension	Language development Individual differences	*Lukas hat {*dem/ keinem} … **{so recht/ jemals}**…* Lukas has {*the/ no} … {**really**$_{NPI}$/ **ever**$_{NPI}$}… *Lukas hat {dem/ #keinem} … **{durchaus/ absolut}**…* Lukas has {the/ #no} … {**quite**$_{PPI}$/ **absolutely**$_{PPI}$}…

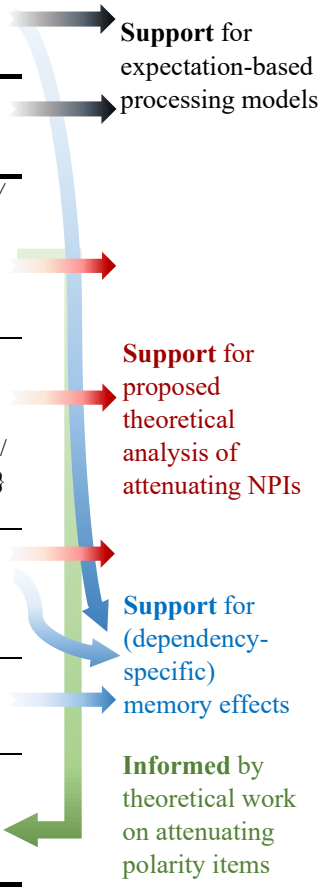


Table I.1: Overview of the included publications and their internal relations.

I.2: EXPECTATION- AND MEMORY-BASED SENTENCE PROCESSING MODELS

Language processing requires rapid processing and integration of incoming verbal (and non-verbal)[3] linguistic input with previously built representations. This is widely accepted to be an incremental process (e.g., Tanenhaus et al., 1995) that is sensitive to local syntactic (e.g., Bader & Häussler, 2009; Franck et al., 2002; Van Dyke, 2007) and semantic (e.g., Novick et al., 2008; Sedivy et al., 1999; Tan et al., 2017; Traxler et al., 1998) constraints, as well as to constraints imposed by the global discourse (Nieuwland & Van Berkum, 2006).

Cognitive language processing models provide a link between empirical observation, e.g., in the form of processing differences between two constructions, and the purported underlying mechanisms given a set of assumptions about (a) the cognitive architecture and (b) the underlying grammar. As such, they are a critical tool to determine testable predictions for specific processing theories, essentially narrowing the empirical hypothesis space with respect to both *where* differences between constructions are presumed to emerge and *what* effect size we should expect, even for constructions that may not have been empirically tested before (Demberg & Keller, 2019).

Some of the most influential accounts proposed in this regard (Gibson, 2000; Levy, 2008; Lewis & Vasishth, 2005) emphasize two key properties that can serve as linking hypothesis between observed word-level processing difficulties and their latent origin: *surprisal* (the negative logarithm of the probability of a word given its context) and *memory retrieval* effects (in the form of decay of to-be-retrieved representation and interference from similar stored representations). In the following, these will be described in detail.

With respect to the former, Levy (2008), building on Hale (2001), argues for an incremental, parallel parser that maintains a set of structural representations compatible with the input received so far, ranked probabilistically in terms of preference given known language statistics and contextual information, and updated with every incoming word. The processing difficulty of a word is determined as the information-theoretic measure of *relative entropy* of the updated probability distribution over possible structures compared to the same probability distribution *prior* to that word (which is

[3] Non-verbal input that is relevant for language processing famously includes a range of visual cues, such as co-speech gestures (e.g., Bernardis et al., 2010; Habets et al., 2011; Holle et al., 2008; Holler et al., 2018; Kelly et al., 2010), facial expressions (Bavelas & Chovil, 2018; Benitez-Quiroz et al., 2016; Carminati & Knoeferle, 2013; Maquate & Knoeferle, 2021), and non-linguistic visual contextual information (e.g., Knoeferle et al., 2008; Novick et al., 2008; Sedivy et al., 1999; Tanenhaus et al., 1995).

mathematically equivalent to *surprisal* for that word; Levy, 2008). Thus, a word's processing difficulty is proportional to the degree to which it reduces uncertainty about the structural representation.

To illustrate the mechanics of this model, we may consider two domains to which it has successful been applied, namely ambiguity resolution effects and effects of (syntactic) expectations. For the first, empirical work on RC attachment (Traxler et al., 1998; van Gompel et al., 2005) has shown an ambiguity advantage for structures like (5b), where *himself* does not resolve the attachment site for the RC (as both *the son* and *the colonel* are masculine target NPs), compared to (5a), in which it does. Levy (2008) demonstrates that this finding is straightforwardly accounted for under surprisal theory, since the conditional probability of *himself* given the prior context will be considerably higher in (5b). (5b) is compatible with both RC attachments considered in parallel by the parser, thus the conditional probability of *himself* is the sum of the probabilities of *himself* under either attachment (6). In (5a), where *himself* is only compatible with low RC attachment, on the other hand, the high attachment parse is eliminated. The second term of the sum in equation (6) thus amounts to zero, resulting, overall, in a lower conditional probability of *himself*.

(5) a. *The daughter$_i$ of the colonel$_j$ who shot himself$_{*i/j}$ on the balcony had been very depressed.*
b. *The son$_i$ of the colonel$_j$ who shot himself$_{i/j}$ on the balcony had been very depressed.*

$$P_i(himself) = P_i(RC_{low})P(himself|RC_{low}) + P_i(RC_{high})P(himself|RC_{high}) \quad (6)$$

(Levy, 2008: 1154)

A second prediction that falls out of the model is that the end constituent in a prospective dependency relation may sometimes be easier to process with an increased number of elements intervening between the two dependents. This is because (a) initial occurrence of the first element in the dependency relation will restrict the structures considered by the parser to those that are compatible with the later appearance of the end element and (b) intervening elements can only further constrain the set of possible structures, which will in turn increase certainty about the location and identity of the end element in the dependency. In effect, the final constituent of the dependency relation will be *less* surprising, resulting in lower processing costs. This prediction, too, has received empirical support from head-final (see **Chapter I.2**) and non-head-final (**Chapter II.1**) dependency relations. Still, particularly for the findings related to head-final constructions, memory-based mechanisms offer an alternative explanation for parts of the empirical data.

Word-level surprisal has also been shown to accurately predict the ERP response measured in the so-called N400 (Frank et al., 2015), a negative deflection occurring between roughly 300-500ms post stimulus onset. The N400 is usually argued to reflect lexical or semantic processing mechanisms (e.g., Lau et al., 2008), as its amplitude is modulated by the lexical-semantic fit between the word at which it is measured and its preceding context. Its relation to surprisal has been used to argue that the N400 may

reflect an update to the probabilistic representation of the sentence (Rabovsky et al., 2018) or the computation of a prediction error between top-down expectations and bottom-up input (Bornkessel-Schlesewsky & Schlesewsky, 2019). In any case, it lends further credence to the relevance of surprisal as a metric in language processing.

What is left unresolved in the original proposal by Levy (2008), is whether (and how) the maintenance of probabilistic sentence representations strains comprehenders' memory resources (but cf. Futrell et al., 2020, discussed below). This, in turn, brings us to a set of models that emphasise limitations in memory as primary constraint on sentence processing, the most prominent members of which are arguably the dependency locality theory (Gibson, 2000) and the cue-based retrieval model (Lewis & Vasishth, 2005). Like surprisal theory, these models admit that comprehenders track incomplete dependencies, such that, at a word initiating a new and incomplete dependency relation, this word is either encoded and stored in memory together with information about its outstanding constituents (Lewis & Vasishth, 2005) or induces a processing cost proportional to the number of constituents that are still needed to complete the dependency relation (Gibson, 2000). Contrary to surprisal theory, however, these models assume that this information, once encoded, will simply be stored in memory until such point that subsequent elements initiate the retrieval of the encoded information. A common assumption in both memory-based models is that the activation level of representations that are stored in memory decays over time, such that re-activating a constituent's representation to the target threshold for retrieval is assumed to be more effortful with increased distance between dependents, resulting in higher observed processing difficulties. Gibson, for instance, proposes to quantify distance-based processing costs by the number of novel discourse referents that are introduced in the material intervening between dependents, presupposing that encoding these referents to memory introduces computational costs that affect the subsequent retrieval of the target representation. Alternative proposals suggest to simply count the number of intervening words (Demberg et al., 2013; Temperley, 2007) or syntactic heads (Alexopoulou & Keller, 2007). As mentioned above, Gibson's dependency locality theory further assumes that there is a storage cost to incomplete (i.e., as yet open) dependencies, such that processing difficulties will be increased with a higher number of open dependencies. The processing cost at the end element of a dependency relation is therefore considered to be a combination of integration costs related to the distance between dependents and storage costs related to the current number of open dependencies, both of which put strain on a limited pool of computational resources.

In contrast to the dependency locality theory, the cue-based retrieval model (Lewis & Vasishth, 2005) proposes to simply formalise decay in terms of the time that has passed since the last retrieval of a word, such that although a word's representation will decay upon initial encoding, each retrieval in subsequent processing operations will trigger a spike in its activation level. A second feature unique to the cue-based retrieval model is the assumption that elements that are stored in memory can exert an influence on processing by virtue of interference at the time of retrieval. The principal idea is that the capacity to actively maintain words in working memory is extremely limited, such

that words will be transferred to a declarative memory component upon initial encoding. The integration of an incoming word with its previous dependents subsequently requires their retrieval from declarative memory, which is guided by so-called "retrieval cues", that is, abstract semantic and syntactic features that serve to identify the target representation. In (7), for instance, the matrix verb *was* is assumed to set retrieval cues for a singular subject as its dependent.[4] In (7a), the target dependent *the reporter* matches both retrieval cues. However, the embedded RC object also matches the singular feature. The presence of this partially matching distractor is argued to cause *inhibitory interference*, i.e., a processing slow-down compared to conditions in which there are no matching distractors. Crucially, interference may also be *facilitatory*. In the ungrammatical sentence (7b), both nouns partially match with the retrieval cues. In such situations, upon initiation of the retrieval request at *was*, both targets provide (partial) positive matches that are evidence for the potential grammaticality of the sentence structure. The parser attempts to resolve the conflict between matching representations by initiating a "race" process, in which both targets compete for sufficient activation levels to be retrieved. The process ends as soon as one of them reaches a threshold level for retrieval. This competitive process, on average, results in faster retrieval times (for either target or distractor) than if there is only a single representation that receives activation (Jäger et al., 2020; Logačev & Vasishth, 2016).

(7) a. *The reporter*{+subject, +singular} *who attacked the senator*{–subject, + singular} ***was*** *widely disliked.*

b. **The reporters*{+subject, –singular} *who attacked the senator*{–subject, +singular} ***was*** *widely disliked.*

Predictions of memory-based accounts have been confirmed in a wide range of constructions. Their shared assumption of activation decay, for instance, is evidenced by locality effects in English embedded RC constructions. For sentence like (8), Grodner and Gibson (2005) found that the embedded verb *sent* is processed more quickly in subject RCs (8a) than object RCs (8b). The straightforward explanation is that retrieval of its dependent *the reporter* is faster in (8a), where it has already been activated at the relative pronoun, than in (8b), where the additional material preceding the embedded verb has led to memory decay. Note also that this finding presents a challenge to surprisal theory, which predicts the opposite pattern (cf. Levy, 2008).

(8) a. *The reporter*$_i$ *who* $\__i$ *sent the photographer to the editor hoped for a story.*

b. *The reporter*$_i$ *who the photographer sent* $\__i$ *to the editor hoped for a story.*

Moreover, similarity-based interference has been attested in a variety of constructions including argument–verb dependencies (Nicenboim et al., 2018; Tan et

[4] Jäger et al. (2017) and Engelmann et al. (2019) discuss the difference between agreement dependencies and non-agreement dependencies in subject-verb dependency relations with respect to cue-based interference effects during verb processing. While both show interference effects (e.g., from distractors matching in number or animacy cues, respectively), the observed patterns are distinct, underscoring that agreement and non-agreement dependencies require different treatments.

al., 2017; Van Dyke, 2007; Van Dyke & McElree, 2006), reflexive binding (Dillon et al., 2013; Jäger et al., 2015; Patil et al., 2016; Xiang et al., 2009), and verb phrase ellipsis (Martin & McElree, 2009; Parker, 2018).

Overall, both surprisal and memory constraints thus appear to affect language processing. However, despite wide-spread acceptance of this view, the interaction between the two mechanisms is still not well-understood (but see the following for advocates of an interactive perspective: Demberg & Keller, 2008; Husain et al., 2014; R. P. Levy & Keller, 2013; Vasishth & Drenhaus, 2011). Recent revised versions of both surprisal theory and the cue-based retrieval model (Engelmann et al., 2019; Futrell et al., 2020) aim to take into consideration some of the empirical findings that have proved challenging for the original models, partially bringing us closer to an integration of both perspectives.

For one, Futrell et al. (2020) propose a *lossy-context* surprisal theory, which maintains most of the original assumptions of surprisal theory, but formalises surprisal in terms of a word's negative logarithmic probability given *a lossy representation of its context*. In essence, the addition of a noise component to the representation of the context allows for the unification of expectation and memory effects within a single model. While the exact noise component one assumes is a free parameter of the model, Futrell et al. have shown that they can capture a range of memory-based findings, including structural forgetting (Frank et al., 2016; Frank & Ernst, 2019; Gibson & Thomas, 1999; Vasishth et al., 2010) and the aforementioned locality effects, by assuming a progressive erasure noise according to which words stored in memory will be deleted with some progressive probability *p(e)*. Under this assumption, locality effects like in (8) are a consequence of changes to the memory representation such that the dependent element (here, the verb) can no longer be predicted from the degraded representation at hand.

Secondly, with respect to cue-based retrieval, Engelmann et al. (2019) propose a model that revises oversimplified assumption of the original model in order to provide wider empirical coverage. They argue to take into account items' prominence (in terms of their syntactic or discourse roles) and current activation level when determining retrieval processing difficulties related to memory retrieval. Thus, items are assumed to be more easily accessible in memory if they occupy prominent roles like the sentential subject compared to less prominent ones, such as nouns contained inside a prepositional phrase modifier ('*the man with the hat*'). Moreover, similarity-based interference like in (7) is presumed to be a function of the distractor's activation level relative to the target representation. Finally, Engelmann et al. argue that retrieval cues may become associated with multiple features (to varying degrees), as the regular co-occurrence of cues may result in mutual association. Although these modifications do not yet address the interaction of memory and expectations, they have yielded more accurate model predictions for a range of empirical findings (cf. Engelmann, 2019).

To conclude this section, it is worth noting that since memory constraints and surprisal both capture a range of empirical findings, it may be more important to ask

how these mechanisms act, individually and in interaction, on the relevant representations during language processing than it is to ask which of the two mechanisms is "right". Within my thesis, I therefore investigate dependency relations at various levels of representation, and address both the effects of expectations and memory limitations. Although the models discussed in this section primarily focus on *syntactic* dependency relations, they posit themselves as general cognitive models of language processing. Still, as outlined in the introduction, the manner and extent to which each of these mechanisms is engaged during the processing of dependencies at varying of linguistic representation remains far from clear. In the following chapter, I therefore outline the three types of dependency relations that are considered as part of this thesis, along with the varying demands they place on the language processing system and how these are investigated in the remainder of the thesis.

I.3: THE INVESTIGATED PHENOMENA

At the heart of this thesis lies an empirical investigation of the processing of asymmetric dependency relations at various levels of representation. Therein, each of the linguistic phenomena I consider, by virtue of their unique defining features, contributes insight on a different aspect of processing (and comprehension). In the following, I briefly introduce the three phenomena that are considered in **Parts II** and **III**, with focus on their shared and distinct properties related to the processing of dependency relations. Detailed discussions of their respective linguistic properties can be found in **Chapters II.1**, **II.2**, and **III.1**.

In the domain of syntax, I investigate dependency relations between German determiners and RCs. As shown in **Chapter I.1**, German bare determiners generally do not require RC modifiers. However, the morphologically complex definite determiner *derjenige*MASC (likewise *diejenige*FEM, *dasjenige*NEUT, roughly 'the one'), composed of the bare determiner *der/die/das* 'the' and the adjectival ending *-jenig*, is ungrammatical without an RC (or PP modifier, cf. Blümel & Liu, 2020) restricting the set of individuals it refers to. Thus, although both (9a) and (9b) require comprehenders to attach the relative clause to its DP host upon encountering it in the linguistic input, the complex determiner in (9b) may in itself generate expectations for the upcoming relative clause.

(9) a. *Anna hat* ***den*** *Film(, der den Oscar gewann,) Freunden empfohlen.*
Anna has the film that the Oscar won friends recommended
'Anna has recommended the movie (that has won the Oscar) to her friends.'

b. *Anna hat* ***denjenigen*** *Film*(, der den Oscar gewann,) Freunden empfohlen.*
Anna has the-one film that the Oscar won friends recommended
'Anna has recommended the one movie *(that has won the Oscar) to her friends.'

Previous work on expectation- and memory-based mechanisms in syntactic parsing has primarily focused on argument–verb dependencies (e.g., Grodner & Gibson, 2005; Husain et al., 2014; Konieczny, 2000; Levy & Keller, 2013). This work has contrasted the predictions of the two processing accounts by modifying the distance between the verb's core arguments (the verbal subject and/or object) and the verbal head (e.g., (10)) (Konieczny, 2000); while the expectation-based account predicts that the processing of the verb will be facilitated with increased distance from its dependents (due to increased certainty about verb location and identity, see **Chapter I.2**), the memory-based account in principle predicts that both the initial verb prediction and the verbal arguments themselves will decay in memory with increased distance between constituents, thus leading to higher processing costs at the verb.

(10) *Er hat die Rose auf den (kleinen [runden]) Tisch gelegt.*
He has the rose on the small round table laid
'He has laid down the rose on the (small [round]) table.'

The predictions of the expectation-based account have initially been confirmed for verb-final languages like German or Hindi (Husain et al., 2014; Konieczny, 2000; Levy & Keller, 2013); however, two particularities of verb–argument dependencies have put this conclusion into question. For one, a key assumption of the memory-based account described above is that representations of the verb or its arguments will be reactivated upon modification (Lewis & Vasishth, 2005; Vasishth & Lewis, 2006). Thus, an intervening verbal modifier as in (10) would trigger re-activation of the verb prediction, which offers an explanation in terms of the memory-based account that is equally adequate to that of the expectation-based model (for details, see **Chapter II.1**). Secondly, in contrast to the determiner–RC relation in (9), verb–argument dependencies lack comparison conditions in which a verb is not required. Most previous findings therefore rest on relative differences in the processing time of the verb, which may appear at different sentence positions depending on the number of intervening words (as in (10)). Such comparisons are problematic as comprehenders generally tend to speed up over the course of a sentence (Ferreira & Henderson, 1993), which may lead to reduced processing times at a late-appearing verb for reasons that are unrelated to verb-specific syntactic expectations (see also Levy & Keller, 2013).

The determiner–RC relation foregoes both of these challenges (a) by virtue of allowing for a direct comparison between bare and complex determiners, only one of which is presumed to generate RC expectations, and (b) since the RC cannot be modified by other elements prior to its occurrence, re-activation of the RC prediction can be excluded. In **Chapter II.1**, we exploit these properties to test the predictions of the expectation- and memory-based accounts. We find evidence for an expectation-based anti-locality effect in the determiner–RC relation, thereby providing a crucial piece of novel empirical support for expectation-based parsing models.

In contrast to syntactic dependency relations, which are usually constrained to the local clause or sentence, discourse-related dependencies operate between propositions, often spanning across several sentences in the global context. (11) serves to illustrate: having uttered the main conceded argument, marked with *true*, the speaker does not immediately provide the second argument of the concessive discourse relation. Instead, they (a) add an expansion on the main issue they are conceding, and (b) continue to repeat, with an increased level of detail, key issues of the main conceded argument (here, related to sub-components of a proposed energy policy). The second argument of the concessive discourse relation, relating to the full conceded argument in (11a), follows three sentences after the initial marking with *true.*

(11) (a) **Conceded argument:** *True, it is useful to have an energy policy that not only is as environmentally friendly as possible but also guarantees the energy supply that is crucial to our development,* **(expansion:)** *something that, in Europe, with all our political independence from non-European*

countries, has so far been a distant prospect. I obviously agree on the emissions reduction targets, just as I do on the efforts to develop renewable energy sources. All of that is, of course, right, and the idea of encouraging people to use less-polluting vehicles seems to be a good one.

(b) **Second argument:** *Is it not the case, though, that the situation will end up as it has in Italy[…]?*

Europarl corpus (Koehn, 2005), sentences 841690-841693

It is widely assumed that comprehenders incrementally build a representation of the unfolding discourse structure (Asher & Lascarides, 2003; Asher & Vieu, 2005; Jasinskaja & Karagjosova, 2021), operating under the general constraints on expectation and memory purported above. At the same time, it is readily apparent that the relative flexibility both in how a discourse relation is marked and in the breadth of text it may span over, presents challenges for either mechanism. On the one hand, although markers like *true* (**Chapter II.2**) or discourse connectives like *even so* (Xiang & Kuperberg, 2015) and *on the one hand* (Scholman et al., 2017) are used by comprehenders to anticipate the upcoming discourse (in terms of content and structure), most discourse relations are not explicitly marked (Asr & Demberg, 2015; Das & Taboada, 2018). Among those that are, only some are marked on the first segment of the two related propositions (such as in *although p, q*) rather than at later positions—at which the cue loses its power to anticipate upcoming discourse (such as in *p, but q* or *p. q, though*). Moreover, beyond discourse connectives and related lexical markers, it is unclear whether and to what extent other cues contribute to the generation of discourse expectations; nor do we know how expectations from multiple sources of information are weighted and integrated to inform discourse expectations (but see **Chapter II.2**).

On the other hand,[5] with respect to memory-based mechanisms, the distance over which discourse relations operate taxes any system that assumes that memory representations decay (Engelmann et al., 2019; Lewis & Vasishth, 2005) or degrade (Futrell et al., 2020) over time. In addition, due to the aforementioned flexibility in marking and structure of ongoing discourse, establishing a discourse relation with previous propositions cannot rely on fixed processing routes such as a cue-based retrieval mechanism that relies on retrievable features encoded in the discourse structure or certain discourse markers and connectives. In particular, discourse relations may be unmarked or the connective itself may be ambiguous (e.g., the connective *and*, which, according to the *Penn Discourse Treebank* (Prasad et al., 2008), can mark at

[5] My use of *on the one hand…on the other hand* in this paragraph is not incidental. Readers may take it as an example on which to experience expectations and memory in discourse processing on a first-hand basis. For some readers, who expertly keep track of the discourse structure, the appearance of *on the other hand* may have been a long-expected conclusion to the contrastive discourse relation initiated by *on the one hand* earlier in the paragraph. For others, expectations for *on the other hand* may have waned by the time of its appearance, therefore triggering a backward search for the initial segment marked by *on the one hand*. I count myself in the latter camp. You may attribute this to poor writing on my behalf.

least the following discourse relations: *expansion*, *temporal*, *contingency*, and *comparison*). Arriving at the intended discourse interpretation may therefore have to rely on a probabilistic inference; Asr and Demberg (2020), for instance, have argued that the meaning of a discourse connective is inferred, in context, based on prior expectations about the relations most likely marked by the connective and the linguistic evidence at hand. Contrary to the syntactic dependency relations discussed above, establishing dependency relations in discourse thus arguably operates on a representational level that may require retrieval of both previous discourse *structure* and *meaning* to infer the intended discourse relation—thereby integrating the current proposition into a coherent representation of the ongoing discourse.

In **Chapter II.2**, I examine the cross-linguistic marking of concessive discourse relations in German and English, investigating whether native speakers form discourse expectation for the concessive relation from narrow (lexical) and broad (pragmatic) linguistic cues. Although I find support for facilitating effects of both types of cues (in line with expectation-based mechanisms), these effects act independently and are more stable for the lexical markers. Observed dissociations between on-line processing facilitations localised to individual words and off-line naturalness ratings of the global discourse additionally demonstrate that perceived discourse coherence is (partially) independent of processing ease and draws more heavily on the global pragmatic context than the on-line resolution of discourse relations.

The last phenomenon considered as part of this thesis is situated at the interface of syntax, semantics, and pragmatics, and unites in it some of the key features from both discourse relations and syntactic dependency relations. The phenomenon of polarity sensitivity, representatives of which are so-called positive and negative polarity items (PPIs and NPIs, respectively), refers to the dependency relation between polarity sensitive words or phrases and the contexts they appear in. Therein, NPIs like *ever* or *any* are restricted, broadly speaking, to contexts of negative polarity, including the scope of negation and negation-adjacent (scale-reversing) operators. PPIs like *already*, on the other hand, are repelled by most negative contexts.[6] Questions about what lexical-semantic properties render some expressions polarity sensitive (the *sensitivity question*) and whether their diverse sets of licensing contexts can be unified under a general, shared linguistic property (the *licensing question*) have been discussed extensively in the theoretical literature (among others: Barker, 2018; Chierchia, 2004, 2013; Gajewski, 2011; Giannakidou, 1998, 2006; Israel, 1996, 2011; Kadmon & Landman, 1993; Krifka, 1995; Ladusaw, 1979; Linebarger, 1987). Detailed discussions of these issues can be found in **Chapters III.1**, **III.4**, and **III.5**.

Crucially, one insight from that literature is that the dependency relation between an NPI and its licenser (or licensing context) escapes straightforward explanations in terms of syntactic or semantic feature matching. For one, as demonstrated in (12a), negative

[6] The anti-licensing of PPIs under negation is ripe with exceptions like their acceptability under the scope of metalinguistic negation or double negative constructions (Szabolcsi, 2004). These are discussed in more detail in **Chapter III.5**.

operators like *no* or *not* can only license NPIs from a position in which they take (syntactic or semantic, cf. Ladusaw, 1979) scope over the NPI, such that the sentential negation inside the RC in (12a) cannot license the NPI *ever*. Moreover, as shown in (12b), simply scoping over the NPI may not be sufficient either; although both *every* and *no* are NPI-licensing quantifiers, only *no* can license the NPI *ever* in its scope (Ladusaw, 1979) although both are NPI-licensing in their restrictor, here the prepositional phrase *with any sense*. This has been attributed to the observation that only *no* licenses downward entailing[7] scalar inferences in its scope, such that *no student takes classes in semantics* allows for an inference to alternatives like *no student takes classes in the semantics of polarity items*, whereas the same construction with *every* in place of *no* does not allow for this inference. Lastly, a range of NPI licensers, including *only* (13), *to be surprised*, and *to regret*, are not distinctly negative in their surface form. Still, Horn (2002) has argued that sentences like (13) express two entailments, one positive meaning component (13a) and one exclusive meaning component (13b). The former is the *assertorically inert* component of the utterance, that is, it is backgrounded as entailed, but not asserted part of the utterance. The licensing of NPIs in such sentences occurs by way of the exclusive meaning in (13b), which is the asserted component of the utterance. Similar explanations have been proposed for *to be surprised* and *to regret* (see Giannakidou, 2006; Horn, 2002). Licensing by *only* and related expressions thus in particular demonstrates that the licensing of NPI is an issue related to global properties of the proposition, in this case, which part of the utterance is asserted.[8]

(12) a. **The/No student who had not attended classes ever passed the semantics exam.*
b. **The/*Every/No student with any sense ever takes classes in semantics.*

(13) *Only Mary ever attended any semantics classes.*
a. *Mary attended semantics classes.* (**positive entailment, assertorically inert**)
b. *Noone else other than Mary attended semantics classes.*
(**negative entailment, asserted**)

With respect to the processing of NPIs, the aforementioned aspects have complicated psycholinguistic efforts to characterise the cognitive mechanisms via which comprehenders establish the dependency relation between NPI and licenser (or licensing context). Although a range of works have demonstrated that comprehenders show an immediate sensitivity to the ungrammaticality of NPIs in unlicensed contexts, e.g., by measuring reading times (Parker & Phillips, 2016; Vasishth et al., 2008; Xiang et al., 2013, 2016) or ERP components (Drenhaus et al., 2006; Liu et al., 2019; Saddy et al., 2004; Xiang et al., 2016; Yanilmaz & Drury, 2018; Yurchenko et al., 2013) at the

[7] Downward entailment is defined as such: A function f of type $<\sigma,\tau>$ is downward entailing iff for all x, y of type σ for which $x \Rightarrow y$: $f(y) \Rightarrow f(x)$.

[8] Alternatively, *only* has been argued to license NPIs because it is Strawson downward entailing (von Fintel, 1999), i.e., downward entailing under the condition that the presupposed contents of the involved propositions (here, *Mary takes semantics classes/Mary takes classes in the semantics of polarity items*) are satisfied. For details, see **Chapter III.1**. In any case, under this analysis, too, licensing under *only* depends on inferences about meaning components of the global proposition.

NPI via electroencephalography (EEG), less is known about the computational mechanisms underlying this detection. Valiant efforts to link the processing of NPIs with theoretical assumptions about their licensing property include Szabolcsi et al.'s (2008) attempt to verify whether NPIs facilitate the drawing of downward-entailing inferences, as one may expect if they are licensed by downward entailment. However, the results of that study were inconclusive. Alternatively, experimental work on so-called NPI illusions, where comprehenders temporarily accept NPIs in contexts that contain a licensing negation in a structurally inaccessible position (such as in (12a)), has shown that the illusion may be restricted to a narrow set of intrusive licensers, which offers a novel window into the online computations underlying the licensing of NPIs (Dillon et al., 2013; Muller & Phillips, 2020; Orth et al., 2020; Parker & Phillips, 2016; Vasishth et al., 2008; Xiang et al., 2009, 2013; Yanilmaz & Drury, 2018). However, this work, too, is currently restricted by a focus on a small set of NPIs and licensers, the results for which are compatible with a range of explanations (for details, see **Chapter III.3**).

In **Part III** of this thesis, I approach the topic of polarity sensitivity under a broad scope. Focusing, in particular, on a type of NPI that is called *attenuating* or *understating* (Israel, 1996, 2011),[9] I first address the formal semantic and pragmatic properties that license such NPIs (**Chapter III.1**). Given that background, I then investigate the processing of attenuating NPIs in interaction with two linguistic environments: first, in **Chapter III.2**, I focus on their processing in conditionals—in which pragmatic inferences derived from various types of conditionals show intriguing effects on the licensing of attenuating NPIs. Second, in **Chapter III.3**, I investigate their processing in constructions that are known to give rise to illusory licensing effects—showing that attenuating NPIs do not display illusory licensing from intrusive *no*, which arguable harnesses back to an interaction between their licensing requirements and scalar properties of the intrusive environment. I argue that both findings support the view that the on-line licensing of attenuating NPIs is sensitive to global scalar properties of the linguistic context in which they appear. The findings in **Chapter III.3** furthermore are incompatible with previously proposed cue-based retrieval accounts of NPI licensing. **Chapters III.4** and **III.5** broaden the perspective from processing to comprehension on the basis of two case studies on circumstances in which the comprehension of polarity sensitive expressions falters. **Chapter III.4** investigates German native speakers' comprehension of two attenuating NPIs, showing that, under some circumstances, adults' retrieval of the lexical-semantic features related to the polarity sensitivity of NPIs is subject to interference from form- and meaning-related polarity insensitive competitors, as well as to individual differences in language aptitude. Lastly, **Chapter III.5** investigates NPI and PPI comprehension in 11-12-year-old children, finding that, despite advanced development of their cognitive language faculty, the acquisition of polarity sensitive expressions continues throughout late childhood and

[9] The terms *attenuating NPI* and *understating NPI* are used interchangeably within the publications included as part of this thesis. In the introductory and concluding chapters, I consistently use the term *attenuating NPI*.

adolescence, arguably guided by differences in input frequency and general pragmatic development. Overall, these studies offer support for dependency-specific memory effects for NPI processing, as well as for the crucial role of semantic and pragmatic context properties in licensing polarity sensitive expressions.

Before turning to these studies, the following chapter will introduce the general methodological approach employed across the remainder of the thesis.

I.4: A NOTE ON THE METHODOLOGICAL APPROACH

Overall, within this thesis, I aim to elucidate the mechanisms involved in language processing and comprehension by studying a range of empirical phenomena across two languages (German and English) and varying levels of linguistic representation. In integrating data from various types of dependency relations, I reveal how their unique linguistic properties affect processing and comprehension under the lens of state-of-the-art cognitive language processing models. The defining feature of my approach is that I am combining cutting-edge linguistic theory with psycholinguistic experimentation and advanced statistical modelling. Theoretical linguistic approaches, in particular, are employed to inform predictions about the processing and comprehension of the tested constructions. Psycholinguistic behavioural experiments using a range of methodologies are conducted to test the hypothesis derived from linguistic theory, corpus data, and cognitive language processing models. Finally, Bayesian data analysis techniques are employed to provide a graded assessment of the evidence in line with the tested hypotheses, while taking into account variation related to subject- (e.g., between participants) and group-level (e.g., between children and adults) differences.

Traditionally, the processing and comprehension of language has been addressed using a variety of empirical methods, some on-line (i.e., measures of processing as it occurs) and some off-line (i.e., measures after processing has concluded), in behavioural (e.g., eye-tracking, self-paced reading, judgment data, lexical choice) and neurophysiological (e.g., EEG, functional magnetic resonance imaging) measures (for an overview, see Kaiser, 2013; Schütze & Sprouse, 2013). Therein, each method has potential to be informative about different aspects of processing and comprehension. Reading times elicited by self-paced reading or eye-tracking-while-reading, for instance, can provide a relatively fine-grained assessment of word-level processing difficulties over the course of a sentence; however, these methods are not very informative about the end result of processing, that is, the interpretation that the reader ultimately arrives at. Conversely, judgment data are one of the primary tools for linguistics to determine differences in the grammatical acceptability, semantic well-formedness, or pragmatic felicity of various constructions; yet, these data do not contain any information about whether the measured differences also affect on-line processing of the tested constructions. Within this thesis, I therefore employ a multi-method approach, combining several behavioural measures of on-line processing and off-line comprehension, each tailored to the research questions at hand. Details on the used experimental methods are included in the respective chapters of **Parts II** and **III**.

In addition to experimental research, I also consider corpus data on each of the studied phenomena. Language corpora, i.e., searchable data bases of annotated spoken

or written language data, are a useful tool to guide the formation of hypotheses for psycholinguistic experimentation.[10] Corpus data can provide an estimation of the frequency of certain constructions in a language (although the absence of a construction should not be interpreted as evidence against its acceptability in the language). Moreover, longitudinal data, such as provided in child language corpora, has potential to be informative about the time course at which certain constructions arise during development. For this thesis, I consider monolingual text corpora to extract frequency information about the studied linguistic phenomena, parallel cross-linguistic corpora for an examination of discourse markers, and monolingual child language corpora for insight into the acquisition of polarity sensitive expressions during early and middle childhood. The results of these corpus analyses are examined in conjunction with the elicited experimental data, thus providing a broader perspective on the studied linguistic phenomena.

Finally, a methodological corner stone of this thesis is the use of state-of-the-art Bayesian statistical analysis methods (for excellent introductory texts, see Gelman et al., 2014; McElreath, 2018; Nicenboim & Vasishth, 2016; Sorensen et al., 2016). One of the primarily conceptual advantages of Bayesian data analysis methods is that they provide the researcher with a direct, intuitive measure of the question researchers are interested in, namely the probability of a certain hypothesis given the observed data (14). This probability is determined, on the basis of Bayes rule, as the weighted mean of the so-called *likelihood*, i.e., the conditional probability of the observed data given the hypothesis, and the *prior*, i.e., the *a priori* probability of the hypothesis. Note, crucially, that this differs from the probability (or *p-value*) returned by frequentist null hypothesis significance testing (NHST): the *p*-value is the probability of observing the data (or data that are even more extreme) given that the null hypothesis is true. *P*-values are applied in conjunction with a decision procedure (Dienes, 2011; Gelman et al., 2014), e.g., the binary decision rule to reject the null hypothesis if the *p*-value is lower than some threshold, say, $p < 0.05$. Heated debates around the relative advantages of Bayesian data analysis methods and frequentist methods (primarily NHST) have been waged elsewhere (among others: Berger & Sellke, 1987; Dienes, 2011; Gelman, 2008; Gelman & Stern, 2006; Krantz, 1999). Although both analysis methods arguably have advantages, the Bayesian approach is chosen here for its conceptual benefit of providing graded information about the posterior probability of a hypothesis, instead of the binary outcome of NHST.

$$p(hypothesis|data) \propto p(data|hypothesis)\, p(hypothesis) \tag{14}$$

In large sample data, NHST and Bayesian analyses often yield very similar outcomes (Gelman et al., 2014; Nicenboim & Vasishth, 2016), although, as mentioned above, they differ in the interpretation of the test results. In smaller samples, however, some of the appeals to the Bayesian approach become readily apparent: for one, as the posterior probability is dependent on both likelihood and prior, researchers can and

[10] The discussion here is restricted to corpora of contemporary spoken or written language, as historical language corpora are not immediately relevant to the present research.

should use prior information about plausible hypotheses to maximise the impact of small sample studies. The posterior will be skewed more heavily towards the prior distribution if there is not enough data to suggest otherwise. Under non-informative priors (i.e., priors that are compatible with a wide range of observable effects), small samples therefore result in vague estimates of the true parameter. Conversely, however, small sample studies can still be informative about a given hypothesis if the researcher has highly specific prior knowledge (i.e., the prior distributions are narrowly centred on a particular hypothesis) that can be tested against the novel observed data. Secondly, in relation to the previous issue, the Bayesian approach allows researchers to abandon the dichotomous decision process associated with the *p*-value. As such, instead of accepting or rejecting a hypothesis, researchers can rely on posterior estimates about the most plausible parameter values to provide a graded assessment of the *amount of support* for a particular hypothesis. This is possible even in small sample studies, although in that case a high degree of uncertainty about the posterior parameter values will be retained (Gelman et al., 2014; Nicenboim & Vasishth, 2016).

Finally, in analogy to classical NHST, the Bayesian approach still allows for the computation of a single metric intended to communicate information about the evidence in favour of the researchers' hypothesis. The so-called *Bayes factor* is calculated as the ratio likelihood between the probability of the data given the null hypothesis (H_0) and the alternative hypothesis (H_1) (15). As such, it provides a direct estimate of how much evidence the data provides in favour of one hypothesis over the other (Rouder et al., 2018; Schad et al., 2021; Wagenmakers et al., 2010).

$$Bayes\ factor\ (BF_{01}) = \frac{p(data|H_0)}{p(data|H_1)} \tag{15}$$

The approach pursued in the remainder of this thesis is one that employs Bayesian regression models with non- or weakly informative priors (that is, priors that only constrain parameters to physically plausible values) for the statistical analyses of behavioural data collected using a range of methodologies.[11] The test outcomes are assessed for their graded support of the specified hypotheses, indicated by the mean estimated parameter values and the width of the posterior probability distributions. Where appropriate, evidence for a hypothesis is also determined in terms of the Bayes factor. Lastly, to encourage data re-analysis and replication attempts, all of my stimulus materials, data, and codes are publically available from my *OSF* profile (https://osf.io/vbx6e/) or the individual repositories linked to in the respective chapters.

[11] The data in **Chapter II.2** were originally analysed in the frequentist NHST framework. A Bayesian re-analysis of the data is available in the appendix, however.

Part II

Expectation and Memory in Sentence Processing

II.1: SYNTACTIC PARSING

Schwab, J., Xiang, M., and Liu, M. (2022). Antilocality effect without head-final dependencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(3), 446–463. https://doi.org/10.1037/xlm0001079

Data repository: https://osf.io/h3qa8/

II.2: DISCOURSE PROCESSING

Schwab, J., and 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. https://doi.org/10.5087/dad.2020.20

Data repository: https://osf.io/ux8de/

Part III

Polarity Items: A Dependency Relation at the Interfaces

III.1: LICENSING ATTENUATING NEGATIVE POLARITY ITEMS

Schwab, J., and Liu, M. (to appear). Attenuating NPIs in indicative and counterfactual conditionals. In: Daniel Gutzmann and Sophie Repp (eds.). *Proceedings of Sinn und Bedeutung 26.*

Data repository: https://osf.io/phydx/

.

III.2: PROCESSING ATTENUATING NEGATIVE POLARITY ITEMS IN CONDITIONALS

Schwab, J., and Liu, M. (2022). Processing attenuating NPIs in indicative and counterfactual conditionals. *Frontiers of Psychology.* 13:894396. https://doi.org/10.3389/fpsyg.2022.894396

Data repository: https://osf.io/6zah8/

III.3: ILLUSORY LICENSING OF NEGATIVE POLARITY ITEMS

Schwab, J. (accepted). Lexical variation in NPI illusions – A case study of German *jemals* ‘ever’ and *so recht* ‘really’. *Glossa Psycholinguistics.*

Data repository:

https://osf.io/s9rt8/

III.4: VARIATION IN NEGATIVE POLARITY ITEM COMPREHENSION

Schwab, J., Mueller, J.L., and Liu, M. (accepted). Dimensions of variation in sentence comprehension: A case study on understating negative polarity items in German. *Linguistische Berichte*.

Data repository: https://osf.io/qrc9u/

III.5: THE ACQUISITION OF POLARITY ITEMS

Schwab, J., Liu, M., and Mueller, J.L. (2021). On the acquisition of polarity items: 11- to 12-year-olds' comprehension of German NPIs and PPIs. *Journal of Psycholinguistic Research, 50*, 1487–1509. https://doi.org/10.1007/s10936-021-09801-3

Data repository: https://osf.io/apgnv/

Part IV

Discussion and Conclusion

IV.1: GENERAL DISCUSSION

The questions posed at the outset of my thesis are (a) to what extent comprehenders would engage memory- and expectation-based language processing mechanisms across levels of representation, and (b) how dependency-specific linguistic properties interplay with general language processing mechanisms to affect processing and comprehension in potentially unique ways. To answer these questions, I study the dependency relations between German determiners and RCs (syntax), between components of the discourse, and between polarity items and their licensing contexts (interface of syntax, semantics, and pragmatics). In summing up the research collected in the previous chapters, **Chapters II.1** and **II.2** test predictions of expectation-based processing accounts in syntactic parsing and discourse processing. Using the self-paced reading methodology on linguistic constructions that had not previously been considered in psycholinguistic evaluations, these studies, individually, reveal novel evidence for expectation-based processing models. In **Chapter II.1**, by investigating the non-head-final determiner–RC relation, I confirm an expectation-based anti-locality effect that is not subsumable under alternative, memory-based explanations. In **Chapter II.2**, by comparing lexical and pragmatic cues towards concessive discourse relations in German and English, I find evidence for independent facilitatory effects on the formation of discourse relations between adjacent propositions, with concurrent cross-linguistic differences in off-line preferences for discourse marking. Collectively, both of these studies directly draw on previous work on expectation-based processing, but exploit unique linguistic properties of the tested constructions to the effect of (a) expanding the empirical landscape on expectations in language processing, while (b) refining previously held assumptions about the engagement of expectation-based processing mechanisms across levels of representation.

In **Chapters III.1-III.5**, I narrow in on a dependency relation at the interface of syntax, semantics, and pragmatics: the idiosyncratic relationship between polarity sensitive expressions (with primary focus on NPIs) and the linguistic properties of the (licensing) contexts in which they appear is first considered from a theoretical perspective, and subsequently investigated in the domains of processing, comprehension, and acquisition. Collectively, these studies contribute to our understanding of the distributional restrictions on polarity items, from a formal perspective as well as in regard to adults' and children's sensitivity to licensing constraints during sentence processing and comprehension. Considered individually, **Chapter III.1** first argues for a novel formal linguistic analysis of attenuating NPIs, based in a scalar approach to polarity sensitivity. **Chapters III.1** and **III.2** provide empirical evidence, from cross-linguistic judgment data, in favour of the proposed analysis—and further reveal how these licensing constraints interact with semantic and

pragmatic properties of conditionals as NPI-licensing environment. **Chapter III.3** then turns to the processing of attenuating and strengthening NPIs in illusory licensing constructions. I argue—supported by empirical data and in line with the theoretical proposal—that the illusion harnesses back to an interaction between scalar properties computed during online processing and the licensing constraints of NPIs. Finally, **Chapters III.4** and **III.5** together address limitations in the comprehension of polarity items: first, I show that native speakers' comprehension of the attenuating NPI *sonderlich* is subject to task- and participant-related variation, such that polarity insensitive lexical competitors and a low quality lexical representations occasionally interfere with retrieval of its NPI-related lexical features. Then, I show that the acquisition of polarity items, including attenuating and strengthening NPIs and PPIs, stretches at least into early adolescence, with variation that may be due both to input-driven differences between polarity items and to linguistic differences between NPIs and PPIs, more general, and those of the strengthening and attenuating subtype, in particular.

Altogether, my thesis contributes to theoretical and psycholinguistics *via* three main factors: (a) the breadth of empirical phenomena considered within and across languages, (b) the integration of cutting-edge linguistic theory with psycholinguistic research on language processing, and (c) the application of state-of-the-art statistical analysis methods, including heightened awareness for individual- and task-related variation within and across data samples. Below, I will address each of these contributions in turn, also emphasising interrelations and limitations.

IV.1.1 Empirical scope

With respect to the empirical phenomena, my broad perspective is motivated at once by the distinct linguistic properties that render each of the considered phenomena informative in their own right, and by their potential to reveal shared mechanisms across phenomena and languages. For one, the presented studies show that expectation-based mechanisms emerge across levels of representation: expectations can be generated from lexical elements (e.g., the determiner *derjenige*, the discourse markers *zwar* 'true' and *true/sure*) and broader pragmatic cues, and can target upcoming syntactic or discourse structure. Preliminary findings from an EEG study not included here further suggest that NPIs, too, generate expectations for downstream licensing contexts (Schwab et al., 2022), and these expectations may wane over the course of a sentence. Secondly, my cross-linguistic theoretical and empirical work on NPI licensing suggests that German and English share a scalar licensing mechanism for attenuating NPIs like *all that* and *sonderlich* ('particularly'). At the same time, I show that there are some dependency-specific effects, particularly in memory effects such as the illusory licensing of NPIs and occasional high tolerance for unlicensed uses of the German NPI *sonderlich*, which arguably emerge due to the unique properties that define the NPI–licenser relation. Because the NPI illusion is more selective than current sentence processing models predict and because these models do not straightforwardly foresee the "forgetting" of

NPI-related lexical features or lexical competition from items stored in long-term memory, the observed effects cannot be neatly subsumed under previous discussions of memory effects in sentence processing. My findings suggest that we may need a model of the cognitive architecture that is more flexible about the kinds of representations that are stored and accessed during sentence processing. One step in that direction would be an explicit account of how global sentential properties such as scalarity and pragmatic meaning contributions *via* presuppositions and implicatures interact with word-by-word sentence processing. It is widely held that these properties are computed on-line, although there is considerable debate about the time course and cognitive effort involved (among others: Bill et al., 2018; Bott et al., 2012; Bott & Noveck, 2004; Breheny et al., 2006, 2013; Chemla & Bott, 2013; Cremers & Chemla, 2014; Grodner et al., 2010; Huang & Snedeker, 2009; Romoli & Schwarz, 2015). The present work also suggests that scalar properties of an embedded clause can interfere with main clause processing, and that pragmatic meaning such as implicatures and presuppositions of conditionals directly affect NPI licensing within the conditional antecedent. At the same time, the format in which these meanings are stored alongside the incremental sentence representation and the mechanisms *via* which they are accessed from memory to affect word-by-word processing difficulty remain underspecified; the development of computational cognitive models of the empirical phenomena studied here thus presents a logical next step for this work.

IV.1.2 Theory meets empiricism

In relation to linguistic theory, I contribute to the field by proposing a novel account of the licensing of attenuating NPIs. This proposal addresses a long-standing gap in the theoretical literature on NPIs, which, with few honourable exceptions (Israel, 1996, 2011; Matsui, 2013; Onea & Sailer, 2013; Solt, 2015; Solt & Wilson, 2021), has given little attention to the systematic ways in which attenuating NPIs violate assumptions of extant accounts of NPI licensing (see **Chapter III.1**). My proposed analysis aims to capture the behaviour of attenuating NPIs in German and English, and receives tentative empirical support from the studies included in **Chapters III.1-III.3**. Further cross-linguistic evaluation, as well as an expansion to other attenuating NPIs, is desirable to validate and potentially revise formal details of the proposed analysis. Nonetheless, the primary postulated difference between attenuating NPIs and the strengthening NPI subtype serve as predictive and as explanatory variable for the reported psycholinguistic findings on the processing and comprehension of NPI-containing sentences in subsequent chapters. As a secondary finding, as the analysis hinges on attenuating NPIs' behaviour in conditionals, the proposal and its experimental validation are revealing about semantic and pragmatic properties of various types of conditionals. Hypothetical indicative conditionals, premise conditionals, and counterfactual conditionals are all shown to differ with respect to the inferences comprehenders draw in the course of their processing, which in turn have immediate consequences on the acceptability of NPIs in the conditional antecedent. My research thus shows how

theoretical linguistics can inform psycholinguistic research, specifically when there is a clear linking hypothesis between theoretical analysis and analysis at the level of cognitive computation or representation. In the case of NPIs, this link is provided by the hypothesis that scalar representations involved in NPI licensing have cognitive reality, that is, comprehenders compute (and rely on) scalar meanings during on-line language processing. Similarly, in the case of the reported expectation-based mechanisms in the processing of the determiner–RC relation in German, my predictions on sentence processing are informed by formal linguistic differences between the studied dependency relation and argument–verb dependencies. The latter is head-final in syntactic terms, whereas the former is not, which has immediate consequences on the predictions generated under the compared cognitive processing models. Overall, my thesis thus contributes to two interfacing fields in linguistics and cognitive science more broadly, the formal and the empirical study of language, while highlighting their interrelations within each of the considered phenomena. The integration of both perspectives, when adequately specified for its intersections, in my view holds further potential to advance both fields, by pushing the limits of contemporary psycholinguistic accounts' generalisability across linguistic phenomena, on the one hand, and by testung the empirical validity of formal linguistic proposals, on the other.

IV.1.3 Methodological limitations and outlook

Finally, methodologically, I draw upon state-of-the-art analysis methods in the form of simulation-based prospective power analyses and Bayesian statistics. The benefits of the Bayesian approach are outlined in **Chapter I.3**; they concern the intuitive interpretation of the test statistic as the (posterior) probability of the researcher's hypothesis, on the one hand, and its graded assessment of the amount of support for the hypothesis, on the other. Here, I would like to additionally emphasise a factor that particularly comes to bear on the studies included in **Chapters III.4** and **III.5**, namely observed variation among participants in an experiment. Although all studies included in this dissertation were designed around group-level analyses, the employed mixed effects Bayesian regression models are able to take into account variation between participants in terms of the subject-specific random effects terms included in the model. The fit random effects parameters can be extracted and visualised, offering a novel perspective on cross-participant variation. One part of this variation arguably represents random measurement noise; another part, however, may reveal systematic differences within a sample that is worth further investigation.

Chapters III.4 and **III.5** both touch on the issue of individual differences, either with respect to adult native speakers' comprehension of the NPI *sonderlich* ('particularly') or with respect to 11-12-year-olds' acquisition of four German polarity sensitive expressions. In the latter study, group-level analyses indicate that, as a whole, 11-12-year-olds show some sensitivity to the distributional restrictions of the tested polarity items, but do not display the strong, categorical distinction between licensed and un-/anti-licensed uses that adults do. Subject-level random effects parameters for

the effects of interest, that is, the comparisons between the ratings provided to licensed and un-/anti-licensed uses, are displayed in **Figures IV.1** (adults) and **IV.2** (11-12-year-olds).[12] They reveal interesting differences between these two samples, in that the adult sample is rather "well-behaved", with consistent response patterns across participants and polarity items. 11-12-year-olds, on the other hand, show much more variation around the group-level estimate. Crucially, for a considerable amount of the participating children, effect estimates are centred around zero, that is, they do not distinguish between licensed and un-/anti-licensed polarity items—often for more than one of the tested expressions. Conversely, another set of children consistently show an effect across all four tested polarity items. Speculatively, the differences visualised here may be indicative of individual differences in the progress of language development at age 11-12. To reveal the origin of these differences, future studies should combine a measure of the comprehension of polarity items with measures of individual differences in cognitive and linguistic development, such as working memory (WM) resources, vocabulary size, and pragmatic skills. Furthermore, additional studies with younger and older age groups, as well as longitudinal studies, promise to close our knowledge gaps on the acquisitional process.

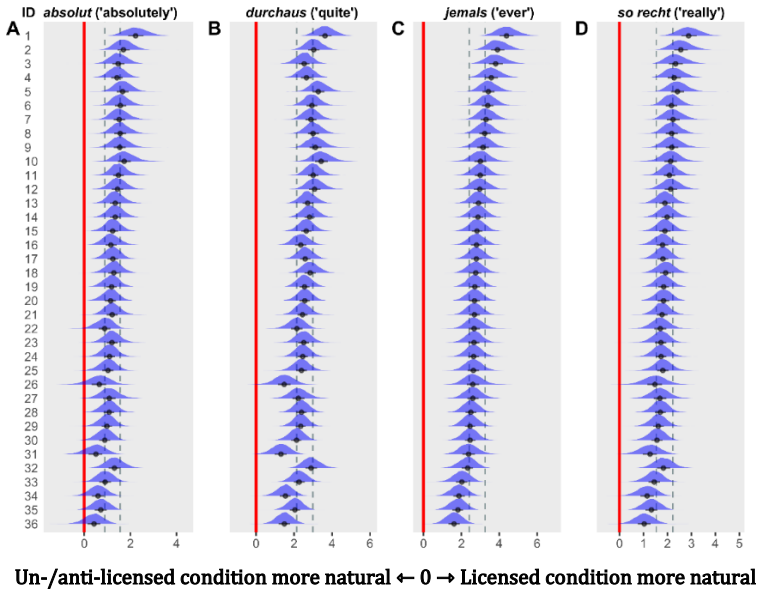


Figure IV.1: Adult participants—Subject-level random effects parameters for the comparison of naturalness ratings provided to licensed and un-/anti-licensed uses of the four polarity sensitive expressions tested in the experiment reported in **Chapter III.5**. The red line indicates the zero intercept (i.e., no effect is present). The dashed lines display the boundaries of the 95% credible interval (CrI) around the group-level parameter estimates. Each blue distribution reflects the posterior probability distribution for one participant, with the solid black lines showing the 50% CrI around the participant's mean estimate marked by a black dot.

[12] Figures can be reproduced from the supplementary analysis scripts included in the *OSF* repository for the paper: https://osf.io/apgnv/

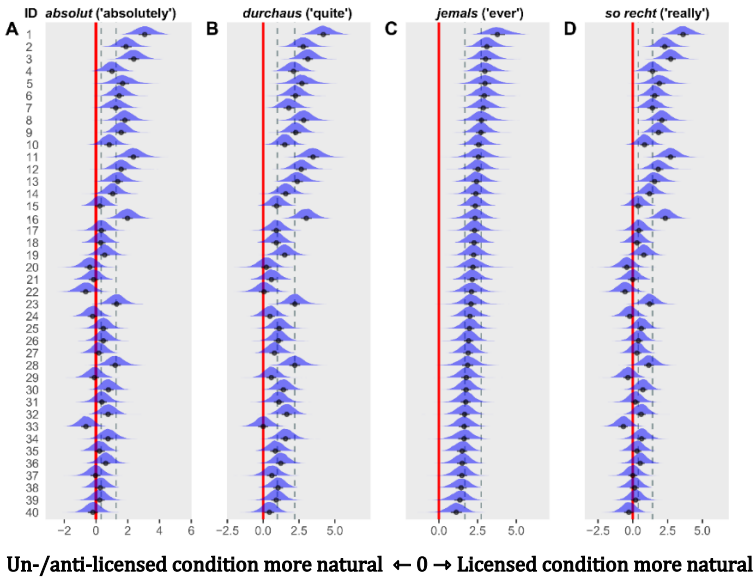


Figure IV.2: 11-12-year-old participants—Subject-level random effects parameters for the comparison of naturalness ratings provided to licensed and un-/anti-licensed uses of the four polarity sensitive expressions tested in the experiment reported in **Chapter III.5**. The red line indicates the zero intercept (i.e., no effect is present). The dashed lines display the boundaries of the 95% credible interval (CrI) around the group-level parameter estimates. Each blue distribution reflects the posterior probability distribution for one participant, with the solid black lines showing the 50% CrI around the participant's mean estimate marked by a black dot.

Similar to the variation observed between children's comprehension of polarity items, **Chapter III.4** is concerned with variation among adult native speakers' comprehension of the NPI *sonderlich*, for which I argue that both task demands and participant-related differences contribute to a high acceptance for unlicensed uses in some contexts. Within the paper, I show that in a self-paced reading and naturalness rating study with complex sentence materials, most participants accept *sonderlich* in positive context to almost the same extent as a polarity insensitive expression, whereas the same does not hold for the NPI *so recht* ('really'). **Figure IV.3**[13], Panels **A** and **B**, show that these effects are relatively consistent across the sample, with most participants showing only a small to non-existent effect for *sonderlich*, but a large effect for *so recht*. In Experiment 3 of the same paper, we remedy these issues by reducing task and stimulus complexity to the point that participants are only asked to provide naturalness ratings for short sentences containing the NPI *sonderlich*, without any time constraint. Here, considerable variation becomes apparent in the subject-level random effects (**Figure IV.3**, Panel **C**). At a group-level, the effect appears as a small dispreference for *sonderlich* in positive (unlicensed) contexts. At the individual level,

[13] Figures can be reproduced from the supplementary analysis scripts included in the *OSF* repository for the paper: https://osf.io/qrc9u/

however, it is apparent that only about half of the participants show an effect in that direction, whereas there is a substantial number of participants for whom it is estimated that there is little to no effect (or a slight tendency in the opposite direction). Suggestively, this may indicate that the reduced cognitive demands of that task are now at level where some participants can succeed at the retrieval of the NPI-related lexical features of *sonderlich*, whereas others still fail. Within the paper, I discuss individual differences in participants' language experience, particularly in the form of vocabulary size and the quality of their lexical representations for an NPI like *sonderlich*, as potential contributing variables to the variation that remains even after task demands have been reduced. Along the same lines as for the acquisition study discussed previously, subsequent work should test these tentative suggestions head-on, for instance by directly manipulating cognitive load in the linguistic task while also assessing individual differences on factors that are suspected to contribute to the variation observed here (e.g., vocabulary size, inhibitory control, and WM resources).

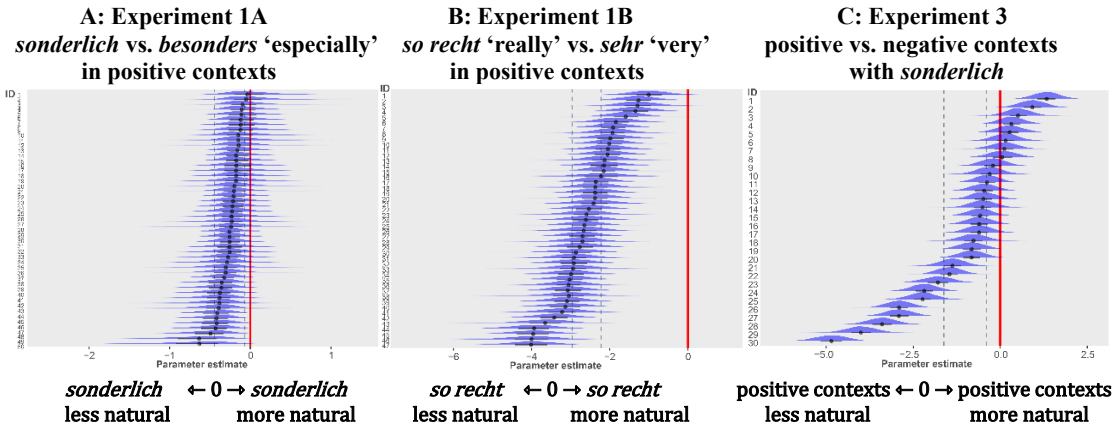


Figure IV.3: Subject-level random effect parameters for the comparison of naturalness ratings provided to (A) affirmative uses of the NPI *sonderlich* and the nonPI *besonders* 'especially', (B) affirmative uses of the NPI *so recht* 'really' and the nonPI *sehr* 'very', as well as to (C) affirmative compared to negative sentences with the NPI *sonderlich*, as reported in Experiments 1A, 1B, and 3 of the paper in **Chapter III.4**. The red line indicates the zero intercept (i.e., no effect is present). The dashed lines display the boundaries of the 95% credible interval (CrI) around the group-level parameter estimates. Each blue distribution reflects the posterior probability distribution for one participant, with the solid black lines showing the 50% CrI around the participant's mean estimate marked by a black dot.

All in all, it is apparent that the limitations of the present work are in the open questions it raises about individual differences in language processing, comprehension, and acquisition. With respect to higher-level language processing,[14] as studied here, individual differences, particularly related to the engagement of predictive mechanisms, have previously been observed between young adults, children, and older adults, with the latter two groups showing less evidence for predictive processing (Federmeier et al., 2010; Gambi et al., 2018; Mani & Huettig, 2012; Payne & Federmeier, 2018; Wlotko & Federmeier, 2012, but cf. Dave et al., 2018). It has been argued that these population differences reflect differences in executive functions, such that individuals with fewer available executive resources (as in children and older adults) are less likely to engage in prediction (e.g., Huettig & Mani, 2016; Pickering & Gambi, 2018). Alternatively, they have also been argued to reflect differences in language experience, such that all populations can engage in prediction, but language experience guides how reliable these predictions are, leading to observable differences in processing measures (e.g., Borovsky et al., 2012; Kaan, 2014)—and populations with consistently unreliable predictions may eventually put greater emphasis on other processing mechanisms (for a review of both hypotheses, see Ryskin et al., 2020). Individual differences have also been reported in syntactic processing, both in children (Woodard et al., 2016) and in younger (James et al., 2018; Swets et al., 2007) and older (Payne et al., 2014) adults. Specifically, syntactic ambiguity resolution has been shown to be subject to individual differences in WM resources and cognitive flexibility, such that for young adults, participants with lower WM resources are biased towards high RC attachment for ambiguous sentences such as *"The maid$_i$ of the princess$_j$ who scratched herself$_{i/j}$ in public was terribly embarrassed"* (James et al., 2018; Swets et al., 2007). This has been argued to be a consequence of "chunking" the word sequences into larger syntactic units (here, the full noun phrase) for more efficient storage. For older adults, on the other hand, this bias has not consistently been confirmed; yet, older adults with higher WM resources spend more time on attempting to resolve the ambiguous attachment than those with fewer WM resources (Payne et al., 2014). For children, ambiguity resolution in garden path sentences has been linked to cognitive flexibility, such that children who perform better at tasks requiring flexible response strategies or switching between different tasks are also more successful in recovering from their initial (incorrect) garden path interpretation (Woodard et al., 2016).

Returning to the topic of polarity items, my own findings demonstrate that this is a phenomenon for which substantial variation between individuals—both young adults and 11-12 year-old children—can be observed. In line with the research reviewed here, future individual difference studies on the comprehension and acquisition of polarity items thus have potential to reveal the language-specific and domain-general cognitive abilities that contribute to the variation observed here; and as the phenomenon is situated at the interface of syntax, semantics, and pragmatics, have simultaneous

[14] Although not relevant to my thesis, individual differences have also been discussed with respect to lower-level, perceptual, processing in language, for instance, in phonological processing (for a review, see Yu & Zellou, 2019).

potential to be revealing about their ways of interaction, and more generally, about the architecture of grammar.

IV.2: CONCLUSION

Dependency relations are ubiquitous in language. Focusing on three asymmetric dependency relations at higher levels of representation (i.e., syntax, discourse, and the interface of syntax, semantics, and pragmatics), I exploit unique linguistic properties of the investigated phenomena to reveal new insight on the cognitive mechanisms involved in their processing and comprehension. By way of example, I thus show that psycholinguistic experimentation and modelling benefit from a broader scope on the phenomena (and languages) they consider, informed, for instance, by cutting-edge linguistic theory. Among the studied phenomena, the dependency relation between polarity items and licensing contexts stand out as one whose properties have attracted substantial debate within both theoretical linguistics and psycholinguistics. To capitalise on respective insights from both fields, I assume a unified approach to one type of polarity item, namely attenuating NPIs, for which I provide a formal analysis, while also taking an empirical perspective on their processing, comprehension, and acquisition. Ultimately, I hold that the integration of both perspectives serves to elevate the insight provided by either, such that each level of analysis can stand in its own right, but research at their interface can act as catalyst for further knowledge gain.

"*When you are a Bear of Very Little Brain, and you Think of Things, you find sometimes that a Thing which seemed very Thingish inside you is quite different when it gets out into the open and has other people looking at it.*"

—Winnie the Pooh, in A.A. Milne's "The House at Pooh Corner"

BIBLIOGRAPHY

Alexopoulou, T., & Keller, F. (2007). *Locality, cyclicity, and resumption: At the interface between the grammar and the human sentence processor*. Language. https://doi.org/10.1353/lan.2007.0001

Aoshima, S., Phillips, C., & Weinberg, A. (2002). Active Filler Effects and Reanalysis: A Study of Japanese. *English*, *12*.

Asher, N., & Lascarides, A. (2003). *Logics of Conversation*. Cambridge University Press.

Asher, N., & Vieu, L. (2005). Subordinating and coordinating discourse relations. *Lingua*, *115*(4 SPEC.ISS.), 591–610. https://doi.org/10.1016/j.lingua.2003.09.017

Asr, F. T., & Demberg, V. (2020). Interpretation of Discourse Connectives Is Probabilistic: Evidence From the Study of But and Although. *Discourse Processes*, *57*(4), 376–399. https://doi.org/10.1080/0163853X.2019.1700760

Asr, F. T., & Demberg, V. (2015). Uniform information density at the level of discourse relations: Negation markers and discourse connective omission. *Proceedings of the 11th International Conference on Computational Semantics*, 118–128.

Badecker, W., & Kuminiak, F. (2007). Morphology, agreement and working memory retrieval in sentence production: Evidence from gender and case in Slovak. *Journal of Memory and Language*, *56*(1), 65–85. https://doi.org/10.1016/j.jml.2006.08.004

Bader, M., & Häussler, J. (2009). Resolving number ambiguities during language comprehension. *Journal of Memory and Language*, *61*(3), 352–373. https://doi.org/10.1016/J.JML.2009.05.005

Barker, C. (2018). Negative polarity as scope marking. *Linguistics and Philosophy*, *41*(5), 483–510. https://doi.org/10.1007/s10988-018-9234-2

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

Bavelas, J., & Chovil, N. (2018). Some pragmatic functions of conversational facial gestures. *Gesture*, *17*(1), 98–127. https://doi.org/10.1075/gest.00012.bav

Benitez-Quiroz, C. F., Wilbur, R. B., & Martinez, A. M. (2016). The not face: A grammaticalization of facial expressions of emotion. *Cognition*, *150*, 77–84. https://doi.org/10.1016/j.cognition.2016.02.004

Berger, J. O., & Sellke, T. (1987). Testing a Point Null Hypothesis: The Irreconcilability of P Values and Evidence. *Journal of the American Statistical Association*, *82*(397), 112–122. https://doi.org/10.1080/01621459.1987.10478397

Bernardis, P., Salillas, E., & Caramelli, N. (2008). Behavioural and neurophysiological evidence of semantic interaction between iconic gestures and words. *Cognitive*

Neuropsychology, *25*(7–8), 1114–1128. https://doi.org/10.1080/02643290801921707

Bill, C., Romoli, J., & Schwarz, F. (2018). Processing Presuppositions and Implicatures: Similarities and Differences. *Frontiers in Communication*, *3*, 44. https://doi.org/10.3389/fcomm.2018.00044

Blake, B. J. (2001). *Case*. Cambridge University Press.

Blümel, A., & Liu, M. (2020). Revisiting obligatory relatives in German – Empirical and theoretical perspectives. *Zeitschrift Für Sprachwissenschaft*, *39*(1), 1–39. https://doi.org/10.1515/zfs-2019-2007

Bock, K., Eberhard, K. M., Cutting, J. C., Meyer, A. S., & Schriefers, H. (2001). Some attractions of verb agreement. *Cognitive Psychology*, *43*(2), 83–128. https://doi.org/10.1006/cogp.2001.0753

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2019). Toward a neurobiologically plausible model of language-related, negative event-related potentials. *Frontiers in Psychology*, *10*(FEB), 298. https://doi.org/10.3389/fpsyg.2019.00298

Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology*, *112*(4), 417–436. https://doi.org/10.1016/j.jecp.2012.01.005

Bott, L., Bailey, T. M., & Grodner, D. (2012). Distinguishing speed from accuracy in scalar implicatures. *Journal of Memory and Language*, *66*(1), 123–142. https://doi.org/10.1016/j.jml.2011.09.005

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, *51*(3), 437–457. https://doi.org/10.1016/j.jml.2004.05.006

Bovolenta, G., & Husband, E. M. (2021). Structural Prediction During Language Comprehension Revealed by Electrophysiology : Evidence from Italian Auxiliaries. *Journal of Experimental Psychology*. https://doi.org/10.1037/XLM0001115

Box, G. E. P., & Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, *26*(2), 211–243. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

Breheny, R., Ferguson, H. J., & Katsos, N. (2013). Investigating the timecourse of accessing conversational implicatures during incremental sentence interpretation. *Language and Cognitive Processes*, *28*(4), 443–467. https://doi.org/10.1080/01690965.2011.649040

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, *100*(3), 434–463. https://doi.org/10.1016/j.cognition.2005.07.003

Bürkner, P.-C., & Vuorre, M. (2019). Ordinal regression models in psychology: A tutorial. *Advances in Methods and Practices in Psychological Science*, *2*(1), 77–101. https://doi.org/10.1177/2515245918823199

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan.

Journal of Statistical Software, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Carminati, M. N., & Knoeferle, P. (2013). Effects of Speaker Emotional Facial Expression and Listener Age on Incremental Sentence Processing. *PLoS ONE*, *8*(9), e72559. https://doi.org/10.1371/journal.pone.0072559

Chemla, E., & Bott, L. (2013). Processing presuppositions: Dynamic semantics vs pragmatic enrichment. *Language and Cognitive Processes*, *28*(3), 241–260. https://doi.org/10.1080/01690965.2011.615221

Chierchia, G. (2004). Scalar implicatures, polarity phenomena and the syntax/pragmatics interface. *Structures and Beyond*.

Chierchia, G. (2006). Broaden your views: Implicatures of domain widening and the "logicality" of language. *Linguistic Inquiry*, *37*(4), 535–590. https://doi.org/10.1162/ling.2006.37.4.535

Chierchia, G. (2013). *Logic in Grammar: Polarity, Free Choice, and Intervention*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199697977.001.0001

Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on Information Theory*, *2*(3), 113–124. https://doi.org/10.1109/TIT.1956.1056813

Chomsky, N. (1957). *Syntactic structures*. Mouton.

Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.

Chomsky, N. (2010). Some simple evo devo theses: How true might they be for language? In *The Evolution of Human Language: Biolinguistic Perspectives* (pp. 45–62). Cambridge University Press. https://doi.org/10.1017/CBO9780511817755.003

Clifton Jr., C., Frazier, L., & Deevy, P. (1999). Feature manipulation in sentence comprehension: 2703. *Rivista Di Linguistica*, *11*(1), 11–40.

Cremers, A., & Chemla, E. (2014). Direct and Indirect Scalar Implicatures Share the Same Processing Signature. In S. P. Reda (Ed.), *Pragmatics, Semantics and the Case of Scalar Implicatures* (pp. 201–227). Palgrave Macmillan. https://doi.org/10.1057/9781137333285_8

Das, D., & Taboada, M. (2018). Signalling of coherence relations in discourse, beyond discourse markers. *Discourse Processes*, *55*(8), 743–770. https://doi.org/10.1080/0163853X.2017.1379327

Dave, S., Brothers, T. A., Traxler, M. J., Ferreira, F., Henderson, J. M., & Swaab, T. Y. (2018). Electrophysiological evidence for preserved primacy of lexical prediction in aging. *Neuropsychologia*, *117*, 135–147. https://doi.org/10.1016/j.neuropsychologia.2018.05.023

Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, *109*(2), 193–210. https://doi.org/10.1016/j.cognition.2008.07.008

Demberg, V., & Keller, F. (2019). Cognitive Models of Syntax and Sentence Processing. In P. Hagoort (Ed.), *Human Language: From Genes and Brains to Behavior* (pp. 293–312). MIT Press.

Demberg, V., Keller, F., & Koller, A. (2013). Incremental, Predictive Parsing with

Psycholinguistically Motivated Tree-Adjoining Grammar. *Computational Linguistics*, *39*(4), 1025–1066. https://doi.org/10.1162/COLI_a_00160

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science*, *6*(3), 274–290. https://doi.org/10.1177/1745691611406920

Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language*, *69*(2), 85–103. https://doi.org/10.1016/J.JML.2013.04.003

Drenhaus, H., Beim Graben, P., Saddy, D., & Frisch, S. (2006). Diagnosis and repair of negative polarity constructions in the light of symbolic resonance analysis. *Brain and Language*, *96*(3), 255–268. https://doi.org/10.1016/j.bandl.2005.05.001

Du, H., Edwards, M. C., & Zhang, Z. (2019). Bayes factor in one-sample tests of means with a sensitivity analysis: A discussion of separate prior distributions. *Behavior Research Methods*, *51*(5), 1998–2021. https://doi.org/10.3758/s13428-019-01262-w

Eberhard, K. M., Cutting, J. C., & Bock, K. (2005). Making syntax of sense: Number agreement in sentence production. *Psychological Review*, *112*(3), 531–559. https://doi.org/10.1037/0033-295X.112.3.531

Engelmann, F., Jäger, L. A., & Vasishth, S. (2019). The Effect of Prominence and Cue Association on Retrieval Processes: A Computational Account. *Cognitive Science*, *43*(12). https://doi.org/10.1111/cogs.12800

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, *115*(3), 149–161. https://doi.org/10.1016/j.bandl.2010.07.006

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15. https://doi.org/10.1111/1467-8721.00158

Ferreira, F., & Henderson, J. M. (1993). Reading processes during syntactic analysis and reanalysis. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *47*(2), 247–275. https://doi.org/10.1037/h0078819

Ferreira, F., & Lowder, M. W. (2016). Prediction, Information Structure, and Good-Enough Language Processing. *Psychology of Learning and Motivation - Advances in Research and Theory*, *65*, 217–247. https://doi.org/10.1016/bs.plm.2016.04.002

Franck, J., Vigliocco, G., & Nicol, J. (2002). Subject-verb agreement errors in French and English: The role of syntactic hierarchy. *Language and Cognitive Processes*, *17*(4), 371–404. https://doi.org/10.1080/01690960143000254

Frank, S. L., & Ernst, P. (2019). Judgements about double-embedded relative clauses differ between languages. *Psychological Research*, *83*(7), 1581–1593. https://doi.org/10.1007/s00426-018-1014-7

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, *140*, 1–11. https://doi.org/10.1016/j.bandl.2014.10.006

Frank, S. L., Trompenaars, T., & Vasishth, S. (2016). Cross-Linguistic Differences in

Processing Double-Embedded Relative Clauses: Working-Memory Constraints or Language Statistics? *Cognitive Science*, *40*(3), 554–578. https://doi.org/10.1111/cogs.12247

Frazier, L., & Flores d'Arcais, G. B. (1989). Filler driven parsing: A study of gap filling in dutch. *Journal of Memory and Language*, *28*(3), 331–344. https://doi.org/10.1016/0749-596X(89)90037-5

Futrell, R., Gibson, E., & Levy, R. P. (2020). Lossy-Context Surprisal: An Information-Theoretic Model of Memory Effects in Sentence Processing. *Cognitive Science*, *44*(3). https://doi.org/10.1111/cogs.12814

Gajewski, J. R. (2011). Licensing strong NPIs. *Natural Language Semantics*, *19*(2), 109–148. https://doi.org/10.1007/s11050-010-9067-1

Gambi, C., Gorrie, F., Pickering, M. J., & Rabagliati, H. (2018). The development of linguistic prediction: Predictions of sound and meaning in 2- to 5-year-olds. *Journal of Experimental Child Psychology*, *173*, 351–370. https://doi.org/10.1016/j.jecp.2018.04.012

Gelman, A. (2008). Objections to Bayesian statistics. *Bayesian Analysis*, *3*(3), 445–450. https://doi.org/10.1214/08-BA318

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian Data Analysis*. CRC Press, Taylor & Francis Group. https://doi.org/10.1201/b16018

Gelman, A., & Stern, H. (2006). The difference between "significant" and "not significant" is not itself statistically significant. *American Statistician*, *60*(4), 328–331. https://doi.org/10.1198/000313006X152649

Giannakidou, A. (1998). Polarity sensitivity as (non) veridical dependency. In *Linguistik Aktuell/Linguistics Today*. John Benjamins Publishing Company. https://doi.org/10.1075/la.23

Giannakidou, A. (2006). Only, emotive factive verbs, and the dual nature of polarity dependency. *Language*, *82*(3), 575–603. https://doi.org/10.1353/lan.2006.0136

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In *Image, language, brain: Papers from the first mind articulation project symposium.* (pp. 94–126). The MIT Press.

Gibson, E., & Thomas, J. (1999). Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes*, *14*(3), 225–248. https://doi.org/10.1080/016909699386293

Goodman, S. N. (1999). Toward Evidence-Based Medical Statistics. 1: The P Value Fallacy. *Annals of Internal Medicine*, *130*(12), 995. https://doi.org/10.7326/0003-4819-130-12-199906150-00008

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentenial complexity. *Cognitive Science*, *29*(2), 261–290. https://doi.org/10.1207/s15516709cog0000_7

Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). "Some," and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, *116*(1), 42–55. https://doi.org/10.1016/j.cognition.2010.03.014

Habets, B., Kita, S., Shao, Z., Özyurek, A., & Hagoort, P. (2011). The role of synchrony and ambiguity in speech-gesture integration during comprehension. *Journal of Cognitive Neuroscience*, *23*(8), 1845–1854. https://doi.org/10.1162/jocn.2010.21462

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. *Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies 2001 - NAACL '01*, 1–8. https://doi.org/10.3115/1073336.1073357

Hammerly, C., Staub, A., & Dillon, B. (2019). The grammaticality asymmetry in agreement attraction reflects response bias: Experimental and modeling evidence. *Cognitive Psychology*, *110*, 70–104. https://doi.org/10.1016/J.COGPSYCH.2019.01.001

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*(5598), 1569–1579. https://doi.org/10.1126/science.298.5598.1569

Holle, H., Gunter, T. C., Rüschemeyer, S. A., Hennenlotter, A., & Iacoboni, M. (2008). Neural correlates of the processing of co-speech gestures. *NeuroImage*, *39*(4), 2010–2024. https://doi.org/10.1016/j.neuroimage.2007.10.055

Holler, J., Kendrick, K. H., & Levinson, S. C. (2018). Processing language in face-to-face conversation: Questions with gestures get faster responses. *Psychonomic Bulletin and Review*, *25*(5), 1900–1908. https://doi.org/10.3758/s13423-017-1363-z

Horn, L. R. (2002). Assertoric intertia and NPI licensing. In M. Andronis, E. Debenport, A. Pycha, & K. Yoshimura (Eds.), *Proceedings of the Thierty-eighth Meeting of the Chicago Linguistic Society* (pp. 55–82). Chicago Linguistic Societa.

Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive Psychology*, *58*(3), 376–415. https://doi.org/10.1016/j.cogpsych.2008.09.001

Huettig, F., & Mani, N. (2016). Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience*, *31*(1), 19–31. https://doi.org/10.1080/23273798.2015.1072223

Husain, S., Vasishth, S., & Srinivasan, N. (2014). Strong expectations cancel locality effects: evidence from Hindi. *PloS One*. https://doi.org/10.1371/journal.pone.0100986

Israel, M. (1996). Polarity sensitivity as lexical semantics. *Linguistics and Philosophy*, *19*(6), 619–666. https://doi.org/10.1007/BF00632710

Israel, M. (2011). *The Grammar of Polarity: Pragmatics, Sensitivity, and the Logic of Scales*. Cambridge University Press. https://doi.org/10.1017/cbo9780511975288

Jäger, L. A., Engelmann, F., & Vasishth, S. (2015). Retrieval interference in reflexive processing: experimental evidence from Mandarin, and computational modeling. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00617

Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language*, *94*, 316–339. https://doi.org/10.1016/j.jml.2017.01.004

Jäger, L. A., Mertzen, D., Van Dyke, J. A., & Vasishth, S. (2020). Interference patterns in subject-verb agreement and reflexives revisited: A large-sample study. *Journal of Memory and Language*, *111*. https://doi.org/10.1016/j.jml.2019.104063

James, A. N., Fraundorf, S. H., Lee, E. K., & Watson, D. G. (2018). Individual differences in syntactic processing: Is there evidence for reader-text interactions? *Journal of Memory and Language*, *102*, 155–181. https://doi.org/10.1016/j.jml.2018.05.006

Jasinskaja, K., & Karagjosova, E. (2021). Rhetorical Relations. In D. Gutzmann, L. Matthewson, C. Meier, H. Rullmann, & T. E. Zimmermann (Eds.), *The Wiley Blackwell Companion to Semantics* (pp. 1–29). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118788516.sem061

Kaan, E. (2014). Predictive sentence processing in L2 and L1. *Linguistic Approaches to Bilingualism*, *4*(2), 257–282. https://doi.org/10.1075/lab.4.2.05kaa

Kadmon, N., & Landman, F. (1993). Any. *Linguistics and Philosophy*, *16*(4), 353–422. https://doi.org/10.1007/BF00985272

Kaiser, E. (2013). Experimental paradigms in psycholinguistics. In R. J. Podesva & D. Sharma (Eds.), *Research Methods in Linguistics* (pp. 135–168). Cambridge University Press.

Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science*, *21*(2), 260–267. https://doi.org/10.1177/0956797609357327

Knoeferle, P., Habets, B., Crocker, M. W., & Münte, T. F. (2008). Visual scenes trigger immediate syntactic reanalysis: Evidence from ERPs during situated spoken comprehension. *Cerebral Cortex*, *18*(4), 789–795. https://doi.org/10.1093/cercor/bhm121

Koehn, P. (2005). Europarl: A parallel corpus for statistical machine translation. *Machine Translation Summit X*, 79–86.

Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, *29*(6), 627–645. https://doi.org/10.1023/A:1026528912821

Konig, E., & Siemund, P. (2000). Causal and concessive clauses: Formal and semantic relations. In E. Couper-Kuhlen & B. Kortmann (Eds.), *Cause - Condition - Concession - Contrast: Cognitive and Discourse Perspectives* (pp. 341–360). Mouton de Gruyter. https://doi.org/10.1515/9783110219043.4.341

Koster, C., & van der Wal, S. (1995). Acquiring a negative polarity verb. *Papers from the German–Dutch Colloquium on Language Acquisition*, 109–126.

Krantz, D. H. (1999). The Null Hypothesis Testing Controversy in Psychology. *Journal of the American Statistical Association*, *94*(448), 1372–1381. https://doi.org/10.1080/01621459.1999.10473888

Krifka, M. (1995). The semantics and pragmatics of polarity items. *Linguistic Analysis*, *25*(3–4), 209–257.

Kuperberg, G. R., Brothers, T., & Wlotko, E. W. (2019). A tale of two positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *Journal of Cognitive Neuroscience*, *32*(1), 12–35. https://doi.org/10.1162/jocn_a_01465

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Ladusaw, W. A. (1979). Polarity sensitivity as inherent scope relations. In *PhD thesis*. University of Texas.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920–933. https://doi.org/10.1038/nrn2532

Lee, M. D., & Wagenmakers, E. J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press. https://doi.org/10.1017/CBO9781139087759

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Levy, R. P., & Keller, F. (2013). Expectation and locality effects in German verb-final structures. *Journal of Memory and Language*, *68*(2), 199–222. https://doi.org/10.1016/j.jml.2012.02.005

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*(3), 375–419. https://doi.org/10.1207/s15516709cog0000_25

Lin, J., Weerman, F., & Zeijlstra, H. (2015). Emerging NPIs: The acquisition of Dutch hoeven "need." *Linguistic Review*, *32*(2), 333–374. https://doi.org/10.1515/tlr-2014-0025

Lin, J., Weerman, F., & Zeijlstra, H. (2018). Acquisition of the Dutch NPI hoeven 'need': From lexical frames to abstract knowledge. *Language Acquisition*, *25*(2), 150–177. https://doi.org/10.1080/10489223.2017.1348097

Linebarger, M. C. (1987). Negative polarity and grammatical representation. *Linguistics and Philosophy*, *10*(3), 325–387. https://doi.org/10.1007/BF00584131

Liu, M. (2019). The elastic nonveridicality property of indicative conditionals. *Linguistics Vanguard*, *5*(s3). https://doi.org/10.1515/lingvan-2019-0007

Liu, M., König, P., & Mueller, J. L. (2019). Novel ERP evidence for processing differences between negative and positive polarity items in German. *Frontiers in Psychology*, *10*, 376. https://doi.org/10.3389/fpsyg.2019.00376

Logačev, P., & Vasishth, S. (2016). A Multiple-Channel Model of Task-Dependent Ambiguity Resolution in Sentence Comprehension. *Cognitive Science*, *40*(2), 266–298. https://doi.org/10.1111/cogs.12228

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake-But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 843–847. https://doi.org/10.1037/a0029284

Maquate, K., & Knoeferle, P. (2021). Integration of Social Context vs. Linguistic Reference During Situated Language Processing. *Frontiers in Psychology*, *12*, 547360. https://doi.org/10.3389/fpsyg.2021.547360

Martin, A. E., & McElree, B. (2009). Memory Operations That Support Language Comprehension: Evidence From Verb-Phrase Ellipsis. *Journal of Experimental Psychology: Learning Memory and Cognition*, *35*(5), 1231–1239.

https://doi.org/10.1037/a0016271

Matsui, A. (2013). Revisiting the licensing problem through understating NPIs – The Case of Japanese anmari '(not) very/much.' In E. Csipak, R. Eckardt, M. Liu, & M. Sailer (Eds.), *Beyond "Any" and "Ever"* (pp. 299–322). De Gruyter Mouton. https://doi.org/10.1515/9783110305234.299

McElreath, R. (2018). Statistical rethinking: A bayesian course with examples in R and stan. In *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. CRC Press. https://doi.org/10.1201/9781315372495

Muller, H., & Phillips, C. (2020). Negative Polarity Illusions. In V. Déprez & M. T. Espinal (Eds.), *The Oxford Handbook of Negation* (pp. 655–676). Oxford University Press. https://doi.org/10.1093/OXFORDHB/9780198830528.013.42

Ness, T., & Meltzer-Asscher, A. (2021). From pre-activation to pre-updating: A threshold mechanism for commitment to strong predictions. *Psychophysiology*, *58*(5), e13797. https://doi.org/10.1111/psyp.13797

Nicenboim, B., & Vasishth, S. (2016). Statistical methods for linguistic research: Foundational Ideas—Part II. *Language and Linguistics Compass*, *10*(11), 591–613. https://doi.org/10.1111/lnc3.12207

Nicenboim, B., Vasishth, S., Engelmann, F., & Suckow, K. (2018). Exploratory and Confirmatory Analyses in Sentence Processing: A Case Study of Number Interference in German. *Cognitive Science*, *42*, 1075–1100. https://doi.org/10.1111/cogs.12589

Nieuwland, M. S., & Van Berkum, J. J. A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, *18*(7), 1098–1111. https://doi.org/10.1162/jocn.2006.18.7.1098

Novick, J. M., Thompson-Schill, S. L., & Trueswell, J. C. (2008). Putting lexical constraints in context into the visual-world paradigm. *Cognition*, *107*(3), 850–903. https://doi.org/10.1016/j.cognition.2007.12.011

Omaki, A., Lau, E., White, I. D., Dakan, M. L., Apple, A., & Phillips, C. (2015). Hyper-active gap filling. *Frontiers in Psychology*, *6*(MAR), 384. https://doi.org/10.3389/fpsyg.2015.00384

Onea, E., & Sailer, M. (2013). Really all that clear? In E. Csipak, R. Eckardt, M. Liu, & M. Sailer (Eds.), *Beyond "Any" and "Ever"* (pp. 323–350). De Gruyter Mouton. https://doi.org/10.1515/9783110305234.323

Orth, W., Yoshida, M., & Sloggett, S. (2020). Negative polarity item (NPI) illusion is a quantification phenomenon. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/XLM0000957

Parker, D. (2018). A memory-based explanation of antecedent-ellipsis mismatches: New insights from computational modeling. *Glossa*, *3*(1). https://doi.org/10.5334/gjgl.621

Parker, D., & Phillips, C. (2016). Negative polarity illusions and the format of hierarchical encodings in memory. *Cognition*, *157*, 321–339. https://doi.org/10.1016/j.cognition.2016.08.016

Patil, U., Vasishth, S., & Lewis, R. L. (2016). Retrieval interference in syntactic processing: The case of reflexive binding in English. *Frontiers in Psychology*, *7*.

https://doi.org/10.3389/fpsyg.2016.00329

Patson, N. D., & Husband, E. M. (2016). Misinterpretations in agreement and agreement attraction. *Quarterly Journal of Experimental Psychology*, *69*(5), 950–971. https://doi.org/10.1080/17470218.2014.992445

Payne, B. R., & Federmeier, K. D. (2018). Contextual constraints on lexico-semantic processing in aging: Evidence from single-word event-related brain potentials. *Brain Research*, *1687*, 117–128. https://doi.org/10.1016/j.brainres.2018.02.021

Payne, B. R., Grison, S., Gao, X., Christianson, K., Morrow, D. G., & Stine-Morrow, E. A. L. (2014). Aging and individual differences in binding during sentence understanding: Evidence from temporary and global syntactic attachment ambiguities. *Cognition*, *130*(2), 157–173. https://doi.org/10.1016/j.cognition.2013.10.005

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement Processes in Sentence Comprehension. *Journal of Memory and Language*, *41*(3), 427–456. https://doi.org/10.1006/JMLA.1999.2653

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002–1044. https://doi.org/10.1037/bul0000158

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., & Webber, B. (2008). The Penn Discourse Treebank 2.0. *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC2008)*, 2961–2968.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rabovsky, M., Hansen, S. S., & McClelland, J. L. (2018). Modelling the N400 brain potential as change in a probabilistic representation of meaning. *Nature Human Behaviour*, *2*(9), 693–705. https://doi.org/10.1038/s41562-018-0406-4

Romoli, J., & Schwarz, F. (2015). An Experimental Comparison Between Presuppositions and Indirect Scalar Implicatures. In F. Schwarz (Ed.), *Experimental Perspectives on Presuppositions. Studies in Theoretical Psycholinguistics* (Vol. 45, pp. 215–240). Springer, Cham. https://doi.org/10.1007/978-3-319-07980-6_10

Rouder, J. N., Haaf, J. M., & Vandekerckhove, J. (2018). Bayesian inference for psychology, part IV: parameter estimation and Bayes factors. *Psychonomic Bulletin and Review*, *25*(1), 102–113. https://doi.org/10.3758/s13423-017-1420-7

Ryskin, R., Levy, R. P., & Fedorenko, E. (2020). Do domain-general executive resources play a role in linguistic prediction? Re-evaluation of the evidence and a path forward. *Neuropsychologia*, *136*. https://doi.org/10.1016/j.neuropsychologia.2019.107258

Saddy, D., Drenhaus, H., & Frisch, S. (2004). Processing polarity items: Contrastive licensing costs. *Brain and Language*, *90*(1–3), 495–502. https://doi.org/10.1016/S0093-934X(03)00470-X

Schad, D. J., Betancourt, M., & Vasishth, S. (2021). Toward a principled Bayesian workflow in cognitive science. *Psychological Methods*, *26*(1), 103–126. https://doi.org/10.1037/met0000275

Scholman, M. C. J., 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. https://doi.org/10.1016/j.jml.2017.07.010

Schütze, C. T., & Sprouse, J. (2013). Judgment data. In R. J. Podesva & D. Sharma (Eds.), *Research Methods in Linguistics* (pp. 27–50). 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. https://doi.org/10.5087/dad.2020.203

Schwab, J., Liu, M., Beese, C., Gruber, T., & Mueller, J. L. (2022). Testing predictive and integrative neural mechanisms in the processing of negative polarity items. *Poster Presented at the 35th Annual Conference on Human Sentence Processing*.

Sedivy, J. C., K. Tanenhaus, M., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*(2), 109–147. https://doi.org/10.1016/S0010-0277(99)00025-6

Solt, S. (2015). Q-adjectives and the semantics of quantity. *Journal of Semantics*, *32*(2), 221–273. https://doi.org/10.1093/jos/fft018

Solt, S., & Wilson, E. C. (2021). M-modifiers, attenuation and polarity sensitivity. In P. G. Grosz, L. Martí, H. Pearson, Y. Soto, & S. Zobel (Eds.), *Proceedings of Sinn und Bedeutung* (Vol. 25, pp. 787–804). https://doi.org/10.18148/SUB/2021.V25I0.967

Sorensen, T., Hohenstein, S., & Vasishth, S. (2016). Bayesian linear mixed models using Stan: A tutorial for psychologists, linguists, and cognitive scientists. *The Quantitative Methods for Psychology*, *12*(3), 175–200. https://doi.org/10.20982/tqmp.12.3.p175

Staub, A. (2009a). On the interpretation of the number attraction effect: Response time evidence. *Journal of Memory and Language*, *60*(2), 308–327. https://doi.org/10.1016/j.jml.2008.11.002

Staub, A. (2009b). On the interpretation of the number attraction effect: Response time evidence. *Journal of Memory and Language*, *60*(2), 308–327. https://doi.org/10.1016/J.JML.2008.11.002

Stowe, L. A. (1986). Parsing wh-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, *1*(3), 227–245. https://doi.org/10.1080/01690968608407062

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, *48*(3), 542–562. https://doi.org/10.1016/S0749-596X(02)00536-3

Swets, B., Desmet, T., Hambrick, D. Z., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General*, *136*(1), 64–81. https://doi.org/10.1037/0096-3445.136.1.64

Szabolcsi, A. (2004). Positive polarity - Negative polarity. *Natural Language and Linguistic Theory*, *22*(2), 409–452. https://doi.org/10.1023/B:NALA.0000015791.00288.43

Szabolcsi, A., Bott, L., & McElree, B. (2008). The Effect of Negative Polarity Items on Inference Verification. *Journal of Semantics*, *25*(4), 411–450. https://doi.org/10.1093/JOS/FFN008

Tan, Y., Martin, R. C., & Van Dyke, J. A. (2017). Semantic and syntactic interference in sentence comprehension: A comparison of working memory models. *Frontiers in Psychology*, *8*(FEB). https://doi.org/10.3389/fpsyg.2017.00198

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634. https://doi.org/10.1126/science.7777863

Temperley, D. (2007). Minimization of dependency length in written English. *Cognition*, *105*(2), 300–333. https://doi.org/10.1016/j.cognition.2006.09.011

Tieu, L., & Lidz, J. (2016). NPI licensing and beyond: Children's knowledge of the semantics of any. *Language Acquisition*, *23*(4), 311–332. https://doi.org/10.1080/10489223.2016.1176172

Traxler, M. J., Pickering, M. J., & Clifton, C. (1998). Adjunct Attachment Is Not a Form of Lexical Ambiguity Resolution. *Journal of Memory and Language*, *39*(4), 558–592. https://doi.org/10.1006/jmla.1998.2600

Van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning Memory and Cognition*, *33*(2), 407–430. https://doi.org/10.1037/0278-7393.33.2.407

Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, *55*(2), 157–166. https://doi.org/10.1016/j.jml.2006.03.007

van Gompel, R. P. G., Pickering, M. J., Pearson, J., & Liversedge, S. P. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language*, *52*(2), 284–307. https://doi.org/10.1016/j.jml.2004.11.003

Vasishth, S., Brussow, S., Lewis, R., & Drenhaus, H. (2008). Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science*, *32*(4), 685–712. https://doi.org/10.1080/03640210802066865

Vasishth, S., & Drenhaus, H. (2011). Locality in German. *Dialogue & Discourse*, *2*(1), 59–82. https://doi.org/10.5087/dad.2011.104

Vasishth, S., & Lewis, R. L. (2006). Argument-Head Distance and Processing Complexity: Explaining both Locality and Antilocality Effects. *Language*, *82*(4), 767–794. https://doi.org/10.1353/lan.2006.0236

Vasishth, S., Mertzen, D., Jäger, L. A., & Gelman, A. (2018). The statistical significance filter leads to overoptimistic expectations of replicability. *Journal of Memory and Language*, *103*, 151–175. https://doi.org/10.1016/J.JML.2018.07.004

Vasishth, S., Suckow, K., Lewis, R. L., & Kern, S. (2010). Short-term forgetting in sentence comprehension: Crosslinguistic evidence from verb-final structures. *Language and Cognitive Processes*, *25*(4), 533–567.

https://doi.org/10.1080/01690960903310587

Vigliocco, G., Hartsuiker, R. J., Jarema, G., & Kolk, H. H. J. (1996). One or more labels on the bottles? Notional concord in Dutch and French. *Language and Cognitive Processes*, *11*(4), 407–442. https://doi.org/10.1080/016909696387169

Vigliocco, G., & Nicol, J. (1998). Separating hierarchical relations and word order in language production: Is proximity concord syntactic or linear? *Cognition*, *68*(1), B13–B29. https://doi.org/10.1016/S0010-0277(98)00041-9

von Fintel, K. (1999). NPI licensing, Strawson entailment, and context dependency. *Journal of Semantics*, *16*(2), 97–148. https://doi.org/10.1093/jos/16.2.97

Wagenmakers, E. J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage-Dickey method. *Cognitive Psychology*, *60*(3), 158–189. https://doi.org/10.1016/j.cogpsych.2009.12.001

Wlotko, E. W., & Federmeier, K. D. (2012). Age-related changes in the impact of contextual strength on multiple aspects of sentence comprehension. *Psychophysiology*, *49*(6), 770–785. https://doi.org/10.1111/j.1469-8986.2012.01366.x

Woodard, K., Pozzan, L., & Trueswell, J. C. (2016). Taking your own path: Individual differences in executive function and language processing skills in child learners. *Journal of Experimental Child Psychology*, *141*, 187–209. https://doi.org/10.1016/j.jecp.2015.08.005

Xiang, M., Dillon, B., & Phillips, C. (2009). Illusory licensing effects across dependency types: ERP evidence. *Brain and Language*, *108*(1), 40–55. https://doi.org/10.1016/j.bandl.2008.10.002

Xiang, M., Grove, J., & Giannakidou, A. (2013). Dependency-dependent interference: NPI interference, agreement attraction, and global pragmatic inferences. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00708

Xiang, M., Grove, J., & Giannakidou, A. (2016). Semantic and pragmatic processes in the comprehension of negation: An event related potential study of negative polarity sensitivity. *Journal of Neurolinguistics*, *38*, 71–88. https://doi.org/10.1016/j.jneuroling.2015.11.001

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

Yanilmaz, A., & Drury, J. E. (2018). Prospective NPI licensing and intrusion in Turkish. *Language, Cognition and Neuroscience*, *33*(1), 111–138. https://doi.org/10.1080/23273798.2017.1371779

Yu, A. C. L., & Zellou, G. (2019). Individual Differences in Language Processing: Phonology. *Annual Review of Linguistics*, *5*, 131–150. https://doi.org/10.1146/annurev-linguistics-011516-033815

Yurchenko, A., den Ouden, D. B., Hoeksema, J., Dragoy, O., Hoeks, J. C. J., & Stowe, L. A. (2013). Processing polarity: ERP evidence for differences between positive and negative polarity. *Neuropsychologia*, *51*(1), 132–141. https://doi.org/10.1016/j.neuropsychologia.2012.10.028

APPENDICES

A.1: BAYESIAN ANALYSIS OF SCHWAB AND LIU (2020)

With the exception of the cross-linguistic self-paced reading and rating study reported in **Chapter II.2**, all behavioural data reported as part of this thesis were analysed within the Bayesian framework to statistical analysis. For methodological consistency, I therefore also report a Bayesian re-analysis of the data in **Chapter II.2**. Details on study methodology and data pre-processing can be found in the relevant publication; the statistical analyses reported below were performed on the same (pre-processed) data as the original frequentist analyses in the paper. All data and code associated with this experiment are available from the *OSF* repository of the associated paper: https://osf.io/ux8de/

1. German experiment

All statistical analyses were carried out using Bayesian regression analyses using the *brms* package, Version 2.12 (Bürkner, 2017) in R, Version 4.0 (R Core Team, 2020). Naturalness ratings were analysed using Bayesian ordinal regression models with a cumulative link function (Bürkner & Vuorre, 2019), whereas reading times (RTs) were analysed using Bayesian linear mixed effects regression models. Just as for the original analyses, the binary predictor variables "contextual cue" (+contextual cue v. -contextual cue) and "lexical cue" (+lexical cue v. -lexical cue) were included as sum-coded fixed effects (with interaction term). The random effects structure was maximal (Barr et al., 2013), that is, it included random by-subject and by-item intercepts and slopes for all fixed effects and their interaction. For the reading time analysis, analyses were performed separately on the critical region (*aber er/sie*, 'but he/she'), as well as the pre- and post-critical regions. I included weakly informative priors for all fixed effects: Priors for the estimated parameters in the rating data were set to a normal distribution with a mean of 0 and a standard deviation of 6. For the reciprocal-transformed RT data, priors were set to a normal distribution with a mean of 0 and a standard deviation of 2. For both rating and RT models, four chains were run with 4,000 sampling iterations each using a warm-up period of 2,000 iterations. I report the posterior parameter estimates together with the 95% credible intervals (CrI) and the posterior probability that the parameter value is bigger/smaller than 0. For all effects on which the posterior distribution suggested an at least 95 % probability of the parameter value being smaller/bigger than 0, Bayes factor analyses were carried out to quantify the amount of evidence for the respective effect. I used the bridge sampling method implemented in the *bayes_factor* function in the R package *brms* to compare a model with the respective effect (H1) to a reduced model in which the effect is removed from

the model (H0). As the Bayes factor is sensitive to the prior distribution (Du et al., 2019; Gelman et al., 2014) and may favour the null hypothesis under uninformative or weakly informative priors (Wagenmakers et al., 2010), the Bayes factor was determined for a range of increasingly informative (that is, more narrowly defined) prior distributions (Jäger et al., 2020; Lee & Wagenmakers, 2013).

1.1 Naturalness ratings

The model was supportive of an effect of contextual cues, $\widehat{\beta} = -0.10, CrI = [-0.19, -0.00], P(\beta < 0) = 0.976$, and also revealed a weak tendency for an effect of lexical cues, $\widehat{\beta} = -0.06, CrI = [-0.14, 0.02], P(\beta < 0) = 0.923$. Both effects were such that the presence of the respective cue increased the perceived naturalness of the stimulus sentence. There was no clear indication of an interaction between these two factors, $\widehat{\beta} = 0.03, CrI = [-0.07, 0.12], P(\beta > 0) = 0.735$.

The Bayes factor analysis on the contextual cue effect revealed clear evidence in favour of the effect under more informative priors, but is equivocal about the effect under less informative priors. This pattern provides evidence in favour of a small effect, while it remains inconclusive about larger effect sizes (cf. Schad et al., 2021).

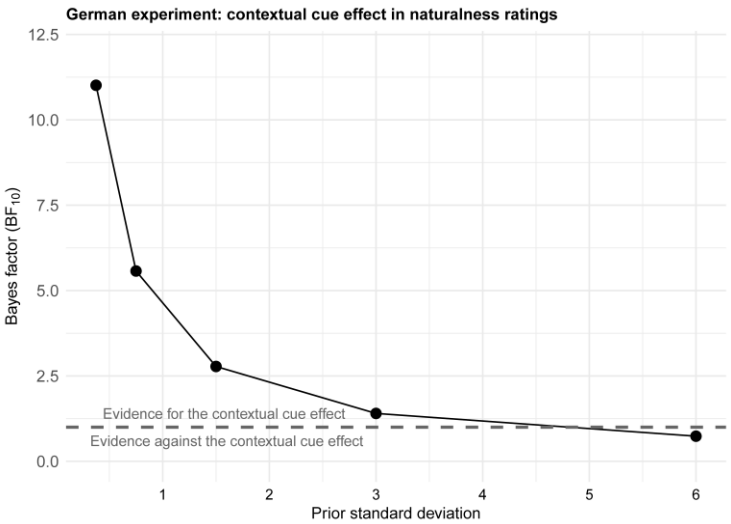


Figure A1: Bayes factor for the contextual cue effect in the naturalness ratings of the German experiment. The plot shows the Bayes factor in favour of the alternative hypothesis versus the null hypothesis (BF_{10}), which was computed for a range of increasingly informative priors.

1.2 Reading times

The model of the critical region provided clear support for a main effect of the lexical cue, $\widehat{\beta} = -0.05, CrI = [-0.08, -0.03], P(\beta < 0) = 0.999$. It was also weakly indicative of a main effect of the contextual cue, $\widehat{\beta} = -0.03, CrI = [-0.05, 0.00], P(\beta < 0) = 0.971$. Both effects were such that the presence of the

respective cue lead to reduced RTs at the critical region. There was no indication for an interaction between these two effects, $\widehat{\beta} = 0.00, CrI = [-0.03, 0.03], P(\beta > 0) = 0.531$. The Bayes factor analysis on the lexical cue effect revealed strong evidence in favour of the effect, regardless of prior informativity (Figure A2). The same analysis on the contextual cue effect, however, favours the null model for most priors and remains equivocal about the contextual cue effect (BF_{10} = 1.04) under even the most informative prior specification (prior *SD* = 0.02).

On the post-critical region, the model was weakly supportive of a continued effect of the lexical cue, $\widehat{\beta} = -0.02, CrI = [-0.05, 0.01], P(\beta > 0) = 0.912$, such that RTs were reduced when the lexical cue was present within the stimulus. There were no indications for effects of the contextual cue or an interaction effect, all $P(\beta < 0) < 0.81$.

On the pre-critical region, there were no indications of any effects, all $P(\beta < 0) < 0.83$.

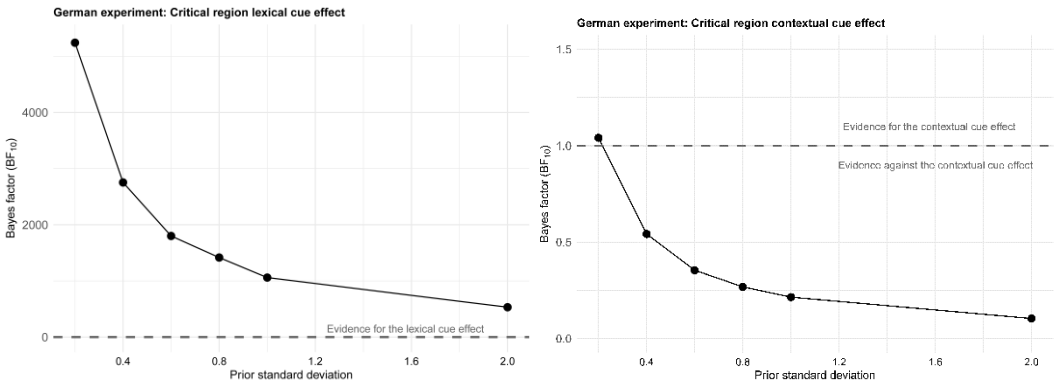


Figure A2: Bayes factor for the lexical cue and contextual cue effects at the critical region of the German experiment. The plot shows the Bayes factor in favour of the alternative hypothesis versus the null hypothesis (BF_{10}), which was computed for a range of increasingly informative priors.

2. English experiment

The model specifications were the same as for the German experiment.

2.1 Naturalness ratings

The model supported an effect of lexical cues, $\widehat{\beta} = 0.19, CrI = [0.09, 0.30], P(\beta > 0) = 0.999$, such that the presence of a lexical cue reduced the naturalness of the sentence. There was no clear indication of an effect of contextual cues, $\widehat{\beta} = -0.04, CrI = [-0.12, 0.03], P(\beta < 0) = 0.867$, but there was a weak tendency for an interaction effect between these two factors, $\widehat{\beta} = -0.05, CrI = [-0.12, 0.01], P(\beta > 0) = 0.950$. The Bayes factor analysis on the lexical cue effect revealed that there was

very strong evidence in favour of the effect, regardless of prior informativity (Figure A3). The same analysis on the interaction effect, however, does not provide evidence in favour of an interaction between contextual and lexical cues.

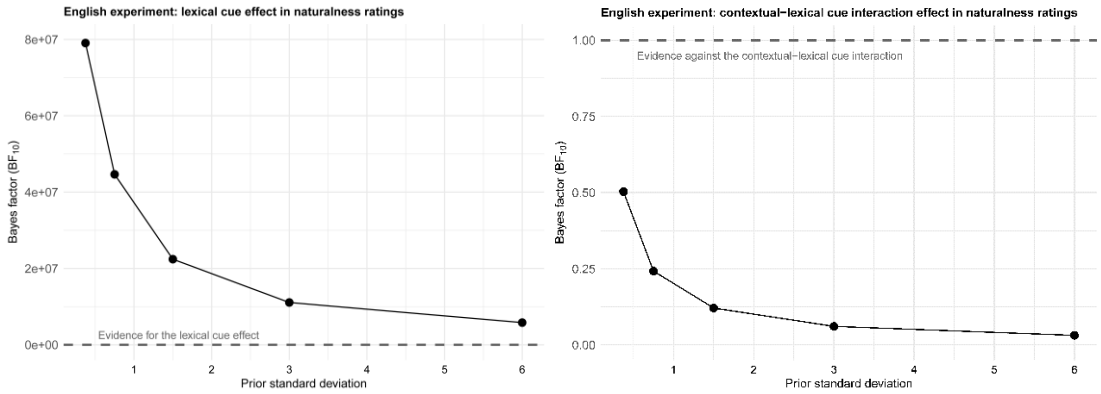


Figure A3: Bayes factor for the lexical cue effect and the interaction between lexical and contextual cue in the naturalness ratings of the English experiment. The plot shows the Bayes factor in favour of the alternative hypothesis versus the null hypothesis (BF_{10}), which was computed for a range of increasingly informative priors.

2.2 Reading times

The model of the critical region revealed no clear indication of any effects, all $P(\beta < 0) < 0.86$. However, on the post-critical region, the model suggested a main effect of lexical cues, such that the region was read faster if a lexical cue had been present within the stimulus, $\widehat{\beta} = -0.04, CrI = [-0.08, -0.01], P(\beta < 0) = 0.989$. There were no clear indication of effects of the contextual cue or an interaction, all $P(\beta < 0) < 0.80$. A Bayes factor analysis on the post-critical lexical cue effect revealed that there was some evidence in favour of the effect, but only under highly informative priors (Figure A4). The remaining uncertainty about the effect warrants future replication.

On the pre-critical region, there was no indication of effects, all $P(\beta < 0) < 0.64$.

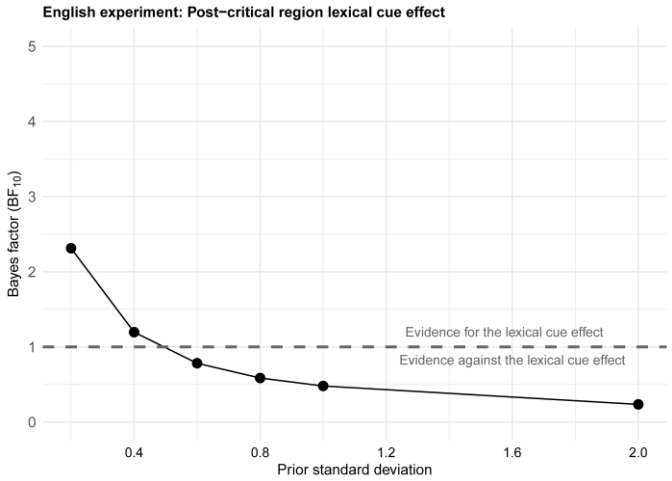


Figure A4. Bayes factor for the lexical cue effect at the post-critical region of the English experiment. The plot shows the Bayes factor in favour of the alternative hypothesis versus the null hypothesis (BF_{10}), which was computed for a range of increasingly informative priors.

3. Discussion

Although the results of the re-analysis are largely in line with the original analysis reported in **Chapter II.2**, they additionally provide a quantificational view onto the evidence in favour of each of the tested effects. This is not possible under traditional Null Hypothesis Significance Testing, both because its test statistic, the *p*-value, is not a measure of the probability of any hypothesis (instead, it represents the probability of the data provided that the null hypothesis was true), and because the *p*-value does not measure size or certainty about the observed effect: a small effect in a big sample can yield the same *p*-value as a big effect in a small sample (Berger & Sellke, 1987; Goodman, 1999). In fact, underpowered small sample studies often over- or under-exaggerate the real underlying effect based on the observed *p*-statistic (Vasishth et al., 2018).

With respect to the naturalness ratings, the Bayesian analysis confirmed that there is strong evidence for the lexical cue effect in the English study and the contextual cue effect in the German study. For interpretations of these effects, please see **Chapter II.2**. The original frequentist analysis, however, had also attested to an interaction between contextual cues and lexical cues in English. The evidence for this effect is weak; even under relatively informative priors, the Bayes factor analysis favours the null model, suggesting that we do not have sufficient evidence to support this effect. In Schwab and Liu (2020: 96), we stated "*the two conditions containing a lexical cue received the lowest numerical ratings […] [T]his may be a reflection of the relatively low frequency of* 'true/sure, ...' *constructions in natural language, as reported in section 3.3.*" Indeed, this is reflected in the lexical cue effect reported above. However, our second conclusion needs to be revised: "*Secondly, we find that the condition containing a contextual cue, but no lexical cue, was rated the most natural. In analogy to the German results, the context may thus have added to the discourse coherence. Further adding a lexical cue to contextually cued items significantly worsened the ratings, suggesting*

that subjects dispreferred lexically cueing a concessive discourse relation if the contextual information already signals a concession." As the present re-analysis does not attest to an effect of contextual cues and does not provide clear evidence in favour of the interaction effect, it remains uncertain whether comprehenders of English concessive constructions rely on narrow contextual cues like the ones we provided to aid their construction of a coherent discourse representations.

In the reading times, too, evidence for the contextual cue effect was not as robust as the evidence for lexical cue effects. Although the original analysis of the critical region effects in the German study attested to effects of both cues (also indicated by their high posterior probability of being smaller than zero in the Bayesian re-analysis), evidence for the contextual cue remains equivocal. Therefore, our original conclusion that "*[o]ur study thus provides evidence that the joint presence of discourse cues from multiple linguistic sources can act as cumulative facilitators of discourse processing.*" (Schwab & Liu, 2020: 95) remains in question. Additional work is needed to verify whether and to what extent comprehenders use contextual cues (in addition or interaction with) lexical cues to aid in discourse processing.

A.2: RT ANALYSIS OMMITTED FROM CHAPTER III.4

Details on study design and methodology can be found in **Chapter III.4**. All analysis reported below were conducted on the same pre-processed data of 47 participants that were analysed with respect to their sentence naturalness ratings in **Chapter III.4**, Experiment 1B. All data and code are available from the *OSF* repository of the associated paper: https://osf.io/qrc9u/.

To identify outliers in the RT data, we first took the whole data set (including fillers) and removed all responses that were more than 2.5 standard deviations from the mean per subject and region, as well as all responses below 150ms (which were most likely erroneous button presses). This affected 3 percent of our data. We then used the *box-cox* procedure (Box & Cox, 1964) to determine the most appropriate transformation to ensure normality of residuals. Following the results of this procedure, we applied a reciprocal transformation on the RT data and multiplied the reciprocal RT by 1000 to make the estimates more interpretable. The latter step does not affect the model results in any way.

All models contained the predictors polarity (NPI, nonPI) and context (affirmative, negative quantifier, sentential negation) as well as their interaction as fixed effects with treatment coded (0, 1) contrasts. We set the priors for the estimated parameters to a normal distribution with mean 0 and standard deviation 3. These priors remain non-committal with regard to the size and direction of the effects, but constrain our prior assumptions to physically plausible effect sizes. For each model, 4 chains were run with 4000 sampling iterations each using a warm-up period of 2000 iterations. The raw reading times are visualised in **Figure A5**; for the analysis we used reciprocal-transformed RTs.

At the first CR (*der/kein Spielplatz*, 'the/no playground'), we did not find support for the predicted interaction between nonPI and NPI and the processing of the definitive vs. negative quantifier, $\widehat{\beta} = 0.03, CrI = [-0.07, 0.13], P(\beta > 0) = 0.722$. Instead, the negatively quantified DP was processed slower than the definite DP for both nonPI $\widehat{\beta} = -0.12\ CrI = [-0.21, -0.03], P(\beta < 0) = 0.997$, and NPI, $\widehat{\beta} = -0.09, CrI = [-0.19, 0.00], P(\beta < 0) = 0.95$.

At the post-critical CR1+1 (*in der Innenstadt*, 'in the city centre'), there were no effects.

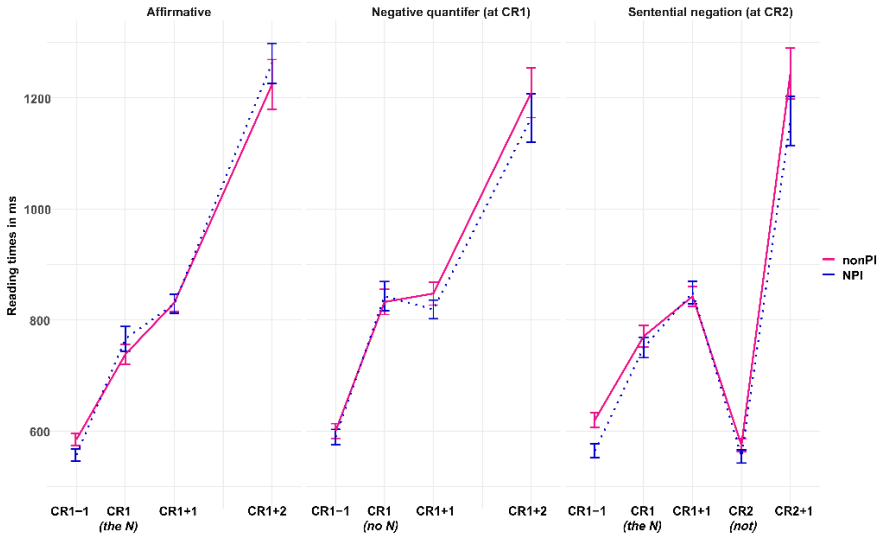


Figure A5. Mean raw RT in milliseconds for the two critical regions of the experiment as well as the immediately preceding and following regions. Error bars show the standard error around the mean.

At the second CR (*nicht*, 'not'), as there are only two conditions we can compare statistically, no interactions are reported. The comparison between the NPI and nonPI condition yielded support for a reduced RT in the NPI condition, $\widehat{\beta} = 0.07, CrI = [0.00, 0.14], P(\beta > 0) = 0.975$, that is, the NPI facilitated processing the negation.

Finally, at the post-critical CR2+1 (*gefallen*, 'liked'), which is also the sentence-final region, there was weak support for an interaction between nonPI and NPI and the comparison between affirmative and negatively quantified sentences, $\widehat{\beta} = 0.09, CrI = [-0.01, 0.19], P(\beta < 0) = 0.961$, as well as strong support for an interaction between nonPI and NPI and the comparison between affirmative sentences and sentences including sentential negation, $\widehat{\beta} = 0.14, CrI = [0.03, 0.25], P(\beta < 0) = 0.995$. Paired comparisons show that the source of these effects lies in an increased RT for the NPI affirmative condition compared to the other conditions, suggesting that unlicensed *so recht* caused a slow-down in processing.

In conclusion, the RT analysis suggests that the presence of an NPI at the sentence beginning can, in principle, facilitate processing negative material downstream. At the sentential negation *nicht* 'not' as well as on the sentence-final region, we found evidence for reduced reading times for the NPI compared to the nonPI condition. However, for the negative quantifier *kein* 'no', we did not find any facilitatory effects on either the quantifier region itself or on subsequent regions. The lack of facilitation with the negative quantifier could harness back to two factors: For one, as the corpus data in **Chapter III.4** shows, *so recht* primarily appears under the scope of sentential negation. The expectations that were generated from *so recht* may thus have placed a high probability on the NPI co-occurring with *nicht*, and only a low probability on the licenser *kein*. In effect, the expectation for a negative quantifier may not have been strong enough to facilitate processing to a measurable degree in our study. The second possibility, not exclusive to the first one, is based in the principle that expectations for

upcoming elements generally become more specific with increased sentence constraint at later sentence positions (Levy, 2008). Thus, comprehenders may not have held very specific expectations about the type and location of licenser within the sentence at the early sentence position at which the negative quantifier appeared. In the latter case, facilitation effects for negative quantifiers could re-emerge if constraints on the predicted licenser position and identity are increased.