Visuo-linguistic integration for situated thematic-role assignment across speakers

Carlotta Isabella Zona

Cumulative Doctoral Dissertation

Submitted to the Faculty of Human Sciences
of the University of Potsdam
in fulfillment of the requirements for the degree
Doctor of Philosophy
in Cognitive Science

2023

Supervisor:
PD Dr. Claudia Felser

Abstract

This dissertation examines the integration of incongruent visual-scene and morphological-case information ("cues") in building thematic-role representations of spoken relative clauses in German.

Addressing the mutual influence of visual and linguistic processing, the Coordinated Interplay Account (CIA) describes a mechanism in two steps supporting visuo-linguistic integration (Knoeferle & Crocker, 2006, *Cog Sci*). However, the outcomes and dynamics of integrating incongruent thematic-role representations from distinct sources have been investigated scarcely. Further, there is evidence that both second-language (L2) and older speakers may rely on non-syntactic cues relatively more than first-language (L1)/young speakers. Yet, the role of visual information for thematic-role comprehension has not been measured in L2 speakers, and only limitedly across the adult lifespan.

Thematically unambiguous canonically ordered (subject-extracted) and noncanonically ordered (object-extracted) spoken relative clauses in German (see 1a-b) were presented in isolation and alongside visual scenes conveying either the same (congruent) or the opposite (incongruent) thematic relations as the sentence did.

1   a   *Das ist der Koch,*   ***der***   *die Braut*   *verfolgt.*

            This is the.NOM cook   who.NOM   the.ACC bride   follows

            This is the cook who is following the bride.

      b   *Das ist der Koch,*   ***den***   *die Braut*   *verfolgt.*

            This is the.NOM cook   whom.ACC   the.NOM bride   follows

            This is the cook whom the bride is following.

The relative contribution of each cue to thematic-role representations was assessed with agent identification. Accuracy and latency data were collected post-sentence from a sample of L1 and L2 speakers (Zona & Felser, 2023), and from a sample of L1 speakers from across the adult lifespan (Zona & Reifegerste, under review). In addition, the moment-by-moment dynamics of thematic-role assignment were investigated with mouse tracking in a young L1 sample (Zona, under review).

The following questions were addressed: (1) How do visual scenes influence thematic-role representations of canonical and noncanonical sentences? (2) How does reliance on visual-scene, case, and word-order cues vary in L1 and L2 speakers? (3) How does reliance on visual-scene, case, and word-order cues change across the lifespan?

The results showed reliable effects of incongruence of visually and linguistically conveyed thematic relations on thematic-role representations. Incongruent (vs. congruent) scenes yielded slower and less accurate responses to agent-identification probes presented post-sentence. The recently inspected agent was considered as the most likely agent ~300ms after trial onset, and the convergence of visual scenes and word order enabled comprehenders to assign thematic roles predictively.

L2 (vs. L1) participants relied more on word order overall. In response to noncanonical clauses presented with incongruent visual scenes, sensitivity to case predicted the size of incongruence effects better than L1-L2 grouping. These results suggest that the individual's ability to exploit specific cues might predict their weighting.

Sensitivity to case was stable throughout the lifespan, while visual effects increased with increasing age and were modulated by individual interference-inhibition levels. Thus, age-related changes in comprehension may stem from stronger reliance on visually (vs. linguistically) conveyed meaning.

These patterns represent evidence for a recent-role preference – i.e., a tendency to re-assign visually conveyed thematic roles to the same referents in temporally coordinated utterances. The findings (i) extend the generalizability of CIA predictions across stimuli, tasks, populations, and measures of interest, (ii) contribute to specifying the outcomes and mechanisms of detecting and indexing incongruent representations within the CIA, and (iii) speak to current efforts to understand the sources of variability in sentence comprehension.

Acknowledgements

Table of Contents

1   General Introduction

One important task in language comprehension is to build THEMATIC-ROLE REPRESENTATIONS – i.e., to understand the relations between sentence referents, or "who is doing what to whom" (Van Gompel & Pickering, 2006). The current work examines the role of information from visual scenes, morphological case, and word order in building such representations.

Research on "situated" sentence processing has considered language as naturally immersed (i.e., situated) in a visual environment (Rosenblum, 2005; 2008). Across the board, visual and linguistic cues are found to be rapidly integrated to constrain thematic-role representations incrementally (i.e., on a constituent-by-constituent basis) and predictively (see Knoeferle, 2019 for a recent review). To account for this evidence, the COORDINATED INTERPLAY ACCOUNT (CIA, Knoeferle & Crocker, 2006) was proposed as a theoretical model describing the temporal mechanisms which underlie the integration of visual and linguistic cues during situated sentence comprehension.

This dissertation assumes the premises of the CIA and extends the evidence supporting it in two ways. Firstly, the current work contributes to defining the "extent and content" (Knoeferle et al., 2014) of thematic-role representations resulting from the integration of competing cues to thematic-role assignment from visual scenes, case, and word order. Secondly, the model's predictions are assessed in second-language speakers and speakers from across the adult lifespan.

To flexibly model competition of word-order and case cues, the stimuli were presented in German. Contrary to English, German makes extensive use of morphological case to map sentence constituents onto the verb's thematic grid, assigning thematic roles regardless of word order (see section 1.1). In manipulating conflict of word-order and case cues, the current work is based on evidence for the SUBJECT ADVANTAGE – i.e., a widely attested preference for canonical, subject-before-object (SO) versus noncanonical, object-before-subject (OS) structures. In German, the subject advantage is indexed by increased processing difficulty and decreased agent-identification accuracy in response to noncanonical OS (vs. canonical SO) structures (Bader & Meng, 1999, 2018, 2023; Meng & Bader, 2021; Schriefers et al., 1995).

Within multiple-constraint models, these findings have typically been attributed to the competition of case and word-order cues to thematic-role assignment of noncanonical sentences, as the subject advantage always endorses a canonical, SO reading, while case cues may convey the opposite interpretation. In addition, SO structures represent the default syntactic structure in most languages (Hammarström, 2016) and are more frequent than OS structures in German (Zubin, 1979). Thus, predictive-processing accounts have also stressed

the role of expectations in facilitating the processing of incoming information that aligns with the currently held interpretation(s).

In what follows, I briefly review constraint-based and predictive accounts of sentence processing, highlighting the relevance of their insights to the thesis' topic ("Models of sentence processing").

Secondly, I turn to visuo-linguistic integration ("Visually situated sentence processing"). The contribution of early models of sentence-picture verification to our current understanding of this integration is highlighted. Then, I introduce the Coordinated Interplay Account as the main theoretical model motivating the thesis' predictions, and I review the evidence supporting it, with a focus on the interaction of visual scenes and the subject advantage in building (predictive) thematic-role representations. Finally, a "recent-role" preference is hypothesized, which this dissertation puts to the empirical test.

Thirdly, I consider evidence on the integration of syntactic and non-syntactic cues for thematic-role assignment in first, and second-language speakers, and in speakers from across the adult lifespan ("Cross-speaker variability in cue weighting"). Importantly, this research suggests that reliance on distinct cue types may be a crucial predictor of cross-speaker variability in sentence comprehension. However, the possible role of visual-scene information as one of the non-syntactic cues shaping thematic-role representations more in second-language (vs. first-language) and older (vs. younger) speakers has not been investigated systematically, so far.

1.1   Morphological case for German thematic-role assignment

In the English sentence *The man is chasing the dog*, the only information necessary to assign thematic roles to the referents is the linear order of their mention. Instead, German uses morphological case to map sentence constituents onto the verb's thematic grid. This is achieved through case markings on pronouns and noun-phrase (NP) determiners (and to some extent via noun suffixes).

Morphological case provides fully contrastive cues for nominative-accusative markings of masculine singular NPs (nominative: *der*, accusative: *den*), while feminine and neuter nominative-accusative markings are identical due to syncretism (nominative: *die/das*, accusative: *die/das*). Contrastive case cues, in turn, may be used to guide thematic-role assignment[1]. For instance, in (2a), the nominative marking on the first determiner (*der*) assigns

---

[1] Note that there is no 1:1 correspondence between case (e.g., nominative) and thematic roles (e.g., agent). For instance, in passive sentences, the nominative case assigns the patient role to the grammatical subject.

the subject grammatical role to the first NP, while the direct-object grammatical role is assigned to the second NP through the determiner in the accusative form (*den*).

2   a   *Der Mann*   *jagt*   *den Hund*

          the.NOM man   chases   the.ACC dog

          The man is chasing the dog.

    b   *Den Mann*   *jagt*   *der Hund*

          the.ACC man   chases   the.NOM dog

          The dog is chasing the man.

This results in a more flexible word order in German over English: While noun-verb-noun (NVN) structures can often be correctly mapped onto subject-verb-object (SVO) structures in English, they may either correspond to SVO or object-verb-subject (OVS) structures in German (Bates & MacWhinney, 1981). For instance, in (2b), the accusative marking on the first determiner (*den*) assigns the direct object role to the first NP, while the nominative marker (*der*) assigns the subject role to the second NP.

Morphological case is also used on relative pronouns to assign a grammatical role to the head noun of relative-clause constructions such as (3a).

3   a   *Maria*   *sieht*   *der Mann,*   **den**   *Mario*   *geschlagen hat*

          Maria   sees   the.NOM man   whom.ACC   Mario   hit has

          The man who Mario hit sees Maria.

Relative-clause constructions are NPs containing a head noun (here, *der Mann*) and a clausal modifier (*den Mario geschlagen hat*). Modifying clauses are often introduced by relativizers signaling a gap (here, *den*). Thus, the head noun bears two grammatical functions, one within the matrix clause (the agent of the seeing action) and one within the relative clause (the patient of the hitting action, Kroeger, 2012).

In the case of feminine and neuter nouns, the syncretism of nominative-accusative case markings may result in situations of global or local thematic ambiguity. For example, the sentence in (4a) is thematically ambiguous at the global level, because no contrastive cues are provided to establish the direction of the action. Instead, the sentence in (4b) is only ambiguous at the local level, as the thematic ambiguity is resolved "by exclusion" through contrastive cues provided at a later stage.

4	a	*Das ist das Kind,*	*das*	*die Frau*	*sucht.*
		This is the.NOM child	who.NOM/ACC	the.NOM/ACC woman	seeks

This is the child who is seeking the woman. / This is the woman who is seeking the child.

	b	*Das ist das Kind*	*das*	*der Mann*	*sucht.*
		This is the.NOM child	who.NOM/ACC	the.NOM man	seeks

This is the child whom the man is seeking.

In cases of thematically ambiguous relative clauses such as (4a), comprehenders often tend to assign the subject grammatical role to the head noun (*das Kind*), as to the first mentioned referent in NVN structures (Bader & Meng, 1999; Schlesewsky & Bornkessel, 2004; Schriefers et al., 1995). For instance, the sentence in (4a) should be overwhelmingly interpreted as meaning *This is the child who is seeking the woman* in virtue of the order of the referents' mention. The sentence in (4b) should initially engender the same expectation (i.e., that the first-mentioned head noun will take on the subject role of the relative clause). Encountering a second-mentioned referent in nominative case disconfirms such expectation, as this assigns the agent role to the second-mentioned referent, implicitly assigning the patient role to the case-ambiguous head noun.

The current thesis exploits such SUBJECT ADVANTAGE – i.e., the preference to interpret the first mentioned referent (in particular, the head noun of relative-clause structures) as the agent of the action – to model the incongruence of word order and morphological case to thematic-role assignment in German.

1.2   Models of sentence processing

1.2.1   *Traditional models*

The time course and outcomes of the integration of cues from different information sources during sentence processing have been the focus of much research in psycholinguistics. This research has traditionally taken one of two main directions.

STRUCTURAL MODELS propose that syntactic structure takes precedence over the processing of non-syntactic cues, which may become relevant only at a later stage (Fodor, 1983; Frazier, 1987; Frazier & Fodor, 1978; Frazier & Rayner, 1982; Friederici, 2011; Gibson, 1998). Instead, CONSTRAINT-BASED MODELS have argued that all relevant information is processed and integrated as soon as it becomes available, regardless of its source (Altmann & Mirković, 2009; Coltheart et al., 2001; Elman et al., 1996; Jurafsky, 1996; McClelland & Elman, 1986; McClelland & Rumelhart, 1981, 1985; Seidenberg & McClelland, 1989). In addition, and contrary to structural models, multiple interpretations are thought to be weighted according to their likelihood (e.g., Altmann & Steedman, 1988). The current dissertation assumes that all

information is extrapolated incrementally (i.e., on a constituent-by-constituent basis) and used to determine the relative likelihood of specific thematic-role configurations between constituents.

The COMPETITION MODEL (CM) is a constraint-based model that has examined how conflicting information from different linguistic cues shapes thematic-role representations by pitting such cues against one another, similar to the current approach. The CM posits that sentences are understood probabilistically through functional relations between form and meaning (Bates et al., 1982). The CM accounts for preferences in thematic-role assignment by positing that form-meaning mappings are weighted differently according to computable dimensions of specific linguistic cues (e.g., case, word order). For example, cue reliability is a dimension calculated as the proportion of times a cue leads to correct thematic-role assignment over its total occurrence. Crucially, the relative weighting of different cues is assumed to vary across languages. For instance, word order leads to correct thematic-role assignment in most English sentences, such that speakers base thematic-role representations mainly on word-order cues (MacWhinney et al., 1984). In contrast, because word order is more flexible and less reliable for German thematic-role assignment, it seems to be relied on to a less extent than in English (Kempe & MacWhinney, 1999; MacWhinney et al., 1984).

However, the design of CM studies typically involved presenting ungrammatical stimuli alongside (semi) grammatical ones (Bates & MacWhinney, 1989). This choice has been criticized for potentially influencing participants' response strategies or causing confusion, thus limiting the conclusions that can be drawn from this evidence (Gibson, 1992). For example, the use of sentences with no meaningful syntactic cues available (e.g., *Licks the cow the goat*, MacWhinney et al., 1984) might have blown the participants' reliance on non-syntactic cues out of proportion.

Rather, using grammatical, unambiguous stimuli represents a more conservative design choice which allows for a parsimonious estimation of the contribution of non-syntactic information to comprehension.

1.2.2   Predictive processing

Research from both structural and constraint-based traditions has often assumed that comprehenders derive meaning first when the cues carrying such meaning are perceived (Ferreira & Chantavarin, 2018). However, on-line methods have revealed that information processing is anticipatory in nature, allowing comprehenders to build upon the currently available cues to infer likely continuations of the input (Spivey & Huette, 2016). This ability, known as *prediction*, has been extensively studied over the past three decades and may be triggered by syntactically and semantically constraining contexts (e.g., Altmann & Kamide, 1999, 2009; Ito et al., 2018; Kamide et al., 2003; Lau et al., 2022; Van Berkum et al., 2005), as

well as by phonological cues, prosody, and intonation (Brown et al., 2011; DeLong et al., 2005; Henry et al., 2017; Weber et al., 2006; Zhang et al., 2021).

For instance, Kamide et al. (2003) (Experiment 2) tracked participants' eye movements as they listened to thematically unambiguous SO/OS sentences in German (see 5a-b).

5	a	*Der Hase*	*frisst*	*gleich*	*den Kohl*
		the.NOM hare	eats	soon	the.ACC cabbage

The hare will soon eat the cabbage.

	b	*Den Hasen*	*frisst*	*gleich*	*der Fuchs*
		the.ACC hare.ACC	eats	soon	the.NOM fox

The fox will soon eat the hare.

In SO conditions, participants directed their fixations towards the depiction of the second noun-phrase (NP) already at the verb. This finding suggests that integrating the semantic and syntactic cues from the first NP and the verb (alongside the visually available entities) allowed participants to identify the upcoming NP prior to its mention. Interestingly, anticipatory fixations were not observed in response to OS structures, indicating that predictive thematic-role assignment was delayed in the face of noncanonical word order.

Indeed, syntactic preferences (such as the subject advantage) may be conceptualized as constraints on prediction. This is because such preferences facilitate the processing of incoming information that aligns with the current interpretation, while encountering an incongruent item leads to parsing difficulty. For example, the difficulty associated with parsing garden-path sentences (e.g., *The horse raced past the barn fell*) is engendered by the expectation of a continuation that fits the initially preferred, simple-clause interpretation (e.g., *The horse raced past the barn quickly*). When this expectation is disconfirmed, the interpretation, and its associated predictions, need to be revised. Similarly, during the resolution of filler-gap dependencies (such as wh-questions and relative clauses), an increased processing difficulty of noncanonical syntactic structures cannot be explained without assuming a predictive mechanism (Kuperberg & Jaeger, 2016). For example, processing object-extracted relative clauses violates the expectations engendered by the subject advantage, as a second NP is unexpectedly encountered after the relativizer (e.g., *the man who Mary*), rather than a finite verb. Supporting this view, syntactically and semantically constraining contexts have been shown to modulate the difficulty associated with processing garden-path and noncanonical sentences (Garnsey et al., 1997; Gordon et al., 2001, 2004; Taraban & McClelland, 1988; Trueswell et al., 1994).

1.2.3   *Dual-route accounts*

Syntactic cues have often been thought to shape sentence representations "algorithmically", i.e., it has often been assumed that only syntactically licensed interpretations are considered. However, evidence has suggested that interpretations at odds with the current syntactic configuration may be derived from highly biasing information, such as semantics and word order. Similar to the current approach, studies in this domain have typically presented thematically unambiguous sentence stimuli in which syntactic and non-syntactic cues conveyed opposite thematic roles, and they have assessed to what extent information from each cue influences thematic-role representations post-sentence with agent-identification questions.

For example, Ferreira (2003) examined thematic-role assignment of English subject and object clefts (e.g., object cleft: *It was the man the dog bit*) while manipulating plausibility (e.g., implausible object cleft: *It was the dog the man bit*). Accuracy was higher for subject clefts overall, and the effect of plausibility (greater accuracy for plausible vs. implausible sentences) was evident in response to object clefts only. These results suggest that the meaning of noncanonical sentences may be inferred from linear word order, and that lexico-semantic cues supporting a specific reading may influence which interpretation is endorsed (Karimi & Ferreira, 2016).

Considering the limited role of non-syntactic information on the processing of syntactically unambiguous sentences, unambiguous case cues to thematic-role assignment have been hypothesized to neutralize or reduce the effects of the subject advantage on agent-identification success. For instance, based on neurophysiological data, the Argument Dependency Model suggests that the presence of unambiguous morphological case marking maximizes the informativeness of case cues, which will lead comprehenders to rely on such cues during comprehension (Schlesewsky & Bornkessel, 2004).

However, pervasive thematic-role misinterpretations have been reported with clauses presenting unambiguous nominative-accusative case cues in German (Bader & Meng, 2018, 2023; Meng & Bader, 2021). Notably, such effects were measured regardless of whether the semantic cues biased towards a syntactically licensed or unlicensed interpretation. For example, Meng & Bader (2021) found significantly decreased accuracy of agent identification in response to unambiguous German OS (vs. SO) sentences regardless of whether the thematic roles were reversible (e.g., *The father hugs the uncle*), biased (e.g., *The chef ruined the roast*), or non-reversible (e.g., *The chef cleaned the pan*). These results suggest a crucial role of linear word order for agent identification in German and showed that unambiguous case markings may not override the subject advantage in thematic-role assignment.

These findings may be interpreted as evidence for a dual-route mechanism underlying comprehension. On the one hand, thematic relations may be understood through the

algorithmic parsing of (morpho) syntactic cues, such as subject-verb agreement and case cues. On the other hand, word-order and lexico-semantic cues may be used as a "fast-and-frugal" heuristic (i.e., a mental short-cut) to understand thematic relations, which is thought to be less demanding than a full syntactic analysis (Ferreira & Patson, 2007). As a potential neural underpinning of dual-route processing, a semantic-memory mechanism has been proposed to be active in parallel to a combinatorial mechanism, indexed by the neurophysiological signatures of processing semantic (N400) and syntactic violations (P600), respectively (Kuperberg, 2007). In Kuperberg's (2007) proposal, such neurophysiological responses are conceptualized as the "error-message" of route-specific predictive streams – yet without assuming a dichotomous division of processing modes.

To summarize, the evidence for misinterpretations of noncanonical sentences indicates a pivotal role of the subject advantage as a heuristic strategy to understanding thematic relations, which may override the use of unambiguous syntactic cues to thematic-role assignment.

1.2.4   Dual-route predictions

Compatible with the evidence for misinterpretations of thematic roles post-sentence, syntactically unlicensed predictions have also been reported in the face of biasing information, challenging the notion of an optimized predictive algorithm guiding sentence processing. Rather, a semantic-priming mechanism has been suggested to be a main driver of predictive behavior (Kuperberg & Jaeger, 2016).

For instance, Kukona et al. (2011) presented sentences such as *Toby arrests the crook* and monitored fixations to depictions of characters which were stereotypically associated with the verb either as the agent or the patient referent (e.g., policeman, crook). At the verb, participants fixated almost equally on stereotypical agent and patient referents – though the agent role had already been assigned. The authors concluded that "local thematic priming can play a highly robust, and potentially larger role in anticipation than sentence-level syntactic constraints" (Kukona et al., 2011, p. 31). Similarly, Kukona et al. (2014) presented spoken sentences containing a thematically constraining verb followed by an adjective (e.g., *The boy will eat the white…*). Participants anticipated not only the upcoming referent (the cake), but also a competitor (a white car) that was incoherent with the verb's thematic restrictions (here, requiring an edible direct object). A similar effect of semantic interference in the ability to integrate local and global constraints in forming predictions was reported by Kamide & Kukona (2018).

The evidence that semantic priming "may play a larger role than syntactic constraints on prediction" (Kukona et al., 2011, p. 31) is compatible with Kuperberg's (2007) view. On the one hand, a semantic memory-based processor may activate adjacent conceptual spaces – potentially giving rise to generic semantic-priming effects. On the other hand, a rule-based

combinatorial processor may enable comprehenders to draw fine-grained predictions on the basis of subtle phonological and morphosyntactic cues, amongst others.

Adding to the latter point, most studies on predictive sentence processing have used the visual-world eye-tracking paradigm (VWP) to infer the time course of the access to linguistic information from the timing of fixations towards visual displays (e.g., Altmann & Kamide, 1999, 2009; Kamide et al., 2003). However, the nature and availability of visual displays has extensively been proven to influence sentence processing (e.g., Sedivy et al., 1999; Tanenhaus et al., 1995) and crucially, enable prediction (e.g., Kamide et al., 2003; Knoeferle & Crocker, 2006). As such, the role of prediction for comprehension may have been systematically overestimated in previous work due to the contribution of visual priming being unaccounted for (Huettig & Mani, 2016). In other words, the displays may simply have provided potential upcoming items "to choose from". The evidence supporting the role of visual information for sentence comprehension, reviewed in the next section, clearly indicates that conclusions from VWP studies may only cautiously be generalized to sentence processing in the absence of the visual component (see also Huettig et al., 2011; Huettig & Mani, 2016; Huettig & McQueen, 2011).

To summarize, there is evidence for a pivotal role of non-syntactic information both on the comprehension of thematic relations, and on the anticipation of upcoming referents. These research findings point to the integration of cues from multiple sources as a main determiner of the likelihood of an interpretation being preferred over another. While evidence for misinterpretations has been provided from manipulations of syntactic, word-order, and semantic cues, visual cues have not been systematically considered by research in this domain.

1.3   Visually situated sentence processing

1.3.1   Early models of sentence-picture verification

Early theoretical models of visuo-linguistic integration, such as the Constituent Comparison Model (CCM, Carpenter & Just, 1975) and the "CACTRAS" model (Clark & Chase, 1972; Trabasso et al., 1971), were specifically proposed to account for response-time patterns of sentence-picture verification (Clark & Chase, 1972; Wannemacher, 1974). In this paradigm, participants respond to questions such as *Is this sentence true of this picture?* Typically, congruent trials are answered faster than incongruent trials. In order to explain this, the CCM and CACTRAS models proposed that both sentences and depictions are understood in terms of their constituents and stored in a propositional format. Then, visually and linguistically derived constituents are thought to be serially verified against each other. Crucially, detecting incongruence of visual and linguistic components triggers the verification process to be re-initiated, adding a constant increment in response times (Carpenter & Just, 1975).

However, the CCM and CACTRAS models have been criticized both on theoretical and methodological grounds (Tanenhaus et al., 1976). In addition, their claims have been challenged by subsequent evidence. For example, verifying incongruent sentence–picture pairs was shown to yield higher accuracy than verifying congruent pairs (Knoeferle et al., 2014). This seems hard to reconcile with the claim that the incongruence of linguistic and visual components triggers the verification process to be re-initiated, as it is unclear how this may benefit verification success. Further, Underwood et al. (2004) found no incongruence effects in sentence–picture verification times when visual scenes and written sentences were presented successively. This suggested that the process of extracting comparable components from visual and linguistic streams for serial comparison may have been associated with the specific settings in previous work (e.g., simultaneous presentation, schematic depictions) and might not generalize to other, more naturalistic settings. For instance, it has been debated whether a verification mechanism underlies situated sentence processing if no explicit comparison is required by the task.

Knoeferle & Crocker (2005) addressed this concern by assessing the effect of visual scenes on thematic-role processing with a task unrelated to sentence-picture congruence, similar to the current approach. Participants saw depictions of two characters as though engaging in a transitive action (e.g., an elderly woman filming a businessman). Then, participants read SO/OS sentences with initial thematic-role ambiguity, disambiguated by case markings on the determiner of the second NP (see 6a-b).

6	a	*Die Oma*	*filmt*	*soeben*	**den** *Handelskaufmann*
		the.NOM/ACC granny	films	currently	the.ACC businessman
		The granny is currently filming the businessman.			
	b	*Die Oma*	*filmt*	*soeben*	**der** *Handelskaufmann*
		the.NOM/ACC granny	films	currently	the.NOM businessman
		The businessman is currently filming the granny.			

Half of the depictions conveyed the same, and half of the depictions conveyed the opposite thematic roles as the sentence. Underwood et al.'s (2004) results were replicated, as the cumulative reading times did not differ as a function of sentence-picture congruence. However, word-by-word reading times exhibited selective slowdowns at the determiner signaling the incongruence, revealing that a comparison was being performed even though the task did not explicitly require it. Furthermore, this pattern suggests that the effects of visuo-linguistic thematic-role incongruence may only be evidenced by fine-grained temporal measures (Knoeferle & Crocker, 2005).

1.3.2   *The Coordinated Interplay Account*

Visual cues have been extensively used in order to infer the timing of linguistic processing with the VWP. The method is based on the LINKING HYPOTHESIS (Cooper, 1974) – i.e., the assumption that linguistic processing influences attention allocation to the visually available entities in a tightly time-locked fashion (Allopenna et al., 1998). In turn, visual information has exhibited a pervasive influence on sentence processing. For instance, Tanenhaus et al. (1995) and Spivey et al. (2002) – amongst others – showed that the difficulty associated with the processing of garden-path instructions (e.g., *Put the apple on the towel in the box*) was determined by the visual display – for example, whether one or two apples were presented visually (see also Sedivy et al., 1999). Since then, ample evidence has been provided supporting the claim that the integration of linguistic and visual cues plays a crucial role in situated sentence processing, and especially in guiding thematic-role assignment (Knoeferle et al., 2005; Knoeferle & Crocker, 2005, 2006)

Accounting for the reciprocal influence of visual information on sentence processing, and of sentence processing on attention allocation, the COORDINATED INTERPLAY ACCOUNT (CIA) was put forth as a theoretical model of visuo-linguistic integration describing a cyclic mechanism in three temporally dependent steps (Knoeferle & Crocker, 2006). Firstly, as utterances are perceived, they are interpreted based on the preceding context and the predictions it engendered. Secondly, the current interpretation directs attention towards referentially relevant entities in the visual environment. Thirdly, the sentence's interpretation and the relevant visual entities are verified against each another. This allows for the grounding of linguistic referents in the real world through their co-indexing to the visually available elements. Crucially, this operation may lead comprehenders to revise the sentence's interpretation based on the visual scene (Knoeferle & Crocker, 2006). The relevant information (e.g., preceding context, visual entities) is stored in working memory and is updated at each cycle (i.e., as the speech unfolds and/or the display changes). Thus, the recency of the visual inspection is thought to play a pivotal role in determining the size of visual effects on comprehension (Crocker et al., 2010).

A crucial question for this dissertation is how visually encoded thematic relations interact with syntactic preferences, such as the subject advantage, in shaping thematic-role representations. Roughly, such interaction has been shown reliably with thematically ambiguous structures, while the evidence from unambiguous structures is more diverse. This pattern broadly aligns with the claim that the influence of non-syntactic cues may be weaker on the comprehension of unambiguous (vs. ambiguous) syntax. In the remainder of this section, I review evidence supporting the CIA, with a focus on the interaction of visually encoded thematic relations and the subject advantage.

In a VWP study, Knoeferle et al. (2005) presented spoken SO/OS structures with initial thematic-role ambiguities (see 7a-b), which were disambiguated by case cues on the determiner of the second NP.

7   a   *Die Prinzessin*   *malt*   *gleich*   **den** *Fechter*

        the.NOM/ACC princess   paints   soon   the.ACC fencer

        The princess will soon paint the fencer.

    b   *Die Prinzessin*   *malt*   *gleich*   **der** *Fechter*

        the.NOM/ACC princess   paints   soon   the.NOM fencer

        The fencer will soon paint the princess.

Fixations were monitored towards displays depicting a character (a princess) as though engaging in unique actions with other characters as an agent or as a patient (e.g., painting a fencer and being washed by a pirate). In response to SO structures, participants fixated on the to-be-mentioned patient upon hearing the verb, revealing that visual and lexico-semantic cues were integrated to understand thematic roles predictively. In response to OS structures, fixations towards the patient and agent characters started diverging only after the onset of the second NP. These results indicate that visually encoded thematic relations guided thematic-role prediction, and that conflicting relations conveyed by visual scenes and word order may delay thematic-role prediction of syntactically ambiguous structures.

Some research has focused on unambiguous structures. For example, Knoeferle & Crocker (2006) (Experiment 2) presented depictions of two referents (e.g., a detective and a wizard) carrying out unique actions onto a patient character (e.g., a pilot) as participants listened to OS structures such as (8a).

8   a   *Den Piloten*   *verzaubert*   *gleich*   *der Detektiv*

        the.ACC pilot.ACC   jinxes   soon   the.NOM detective

        The detective will soon jinx the pilot.

Crucially, the action in each sentence (jinxing) was stereotypically associated with the other character (the wizard), who was carrying out an unrelated action (serving food). This design choice allowed the researchers to tease apart the contribution of two distinct cues to thematic-role processing, similar to the current work. Participants exhibited higher reliance on visual scenes than stereotypical knowledge in assigning the agent role predictively. Furthermore, observing predictive fixations only in response to OS structures suggests that unambiguous case cues may partially neutralize the effects of the subject advantage as a driver

of predictive behavior. These findings further point to the strength of visual effects relative to other non-syntactic sources, such as stereotypical knowledge.

Moreover, Knoeferle (2007) directly compared the effects of visual scenes on thematic-role prediction of thematically ambiguous versus unambiguous SO/OS sentences such as (9a-b).

9   a   ***Die*** *Frau Orange*   *tritt*   *in diesem Moment*   *den Sir Zwiebel.*

the.NOM/ACC Ms. Orange   kicks   in this moment   the.ACC Sir Onion

Ms. Orange is currently kicking Sir Onion.

b   ***Der*** *Herr Orange*   *tritt*   *in diesem Moment*   *den Sir Zwiebel.*

the.NOM Mr. Orange   kicks   in this moment   the.ACC Sir Onion

Mr. Orange is currently kicking Sir Onion.

The visual entities were attended to as soon as the utterance identified them as relevant for predictive thematic-role assignment (i.e., one region earlier for unambiguous than for ambiguous structures). In addition, the patterns indicated that the effects of the subject advantage delayed thematic-role prediction of ambiguous sentences, as the fixations to the upcoming referent were higher than to the first-mentioned character in response to SO, but not OS structures. Conversely, in response to unambiguous sentences, fixations towards the upcoming referent were equal in response to SO and OS structures.

The patterns reported by this research are partially at odds with the results from Kamide et al. (2003), in which thematic-role prediction was not observed in response to unambiguous OS structures. In explaining this discrepancy, it is worth considering that CIA studies have often used action depictions, rather than the depictions of the referents in isolation used by Kamide et al. (2003). Action depictions may have been more immediate to decode semantically than object depictions (for example, because the agent characters are generally depicted as oriented towards the patient) and may have more easily steered visual attention towards the upcoming referents.

Together, these findings support the role of visual-scene cues for thematic-role assignment and specify the time course of the integration of visual, case, word-order cues for thematic-role prediction. Specifically, the results suggest that the presence of unambiguous case cues may neutralize the effects of the subject advantage, which has been shown to delay thematic-role prediction of noncanonical structures.

1.3.3   *The recent-role preference*

The evidence from CIA studies reviewed so far is compatible with a well-oiled predictive mechanism that exploits cues algorithmically to derive meaning. However, research has also

pointed to a RECENT-EVENT PREFERENCE, i.e., the preference to fixate on previously inspected referents as compared to potentially upcoming ones during situated comprehension (e.g., Knoeferle et al., 2011).

For instance, Knoeferle et al. (2011) tracked participants' eye movements as they listened to locally ambiguous spoken sentences which either referred to a recently viewed scene (e.g., *Der Versuchsleiter zuckerte kürzlich die Erbeeren* 'The experimenter recently sugared the strawberries') or to a future action (e.g., *Der Versuchsleiter zuckert demnächst die Pfannkuchen* 'The experimenter will soon sugar the pancakes'). Participants were found to inspect the recent-event object (the strawberries) more than the potentially upcoming referent (the pancakes) regardless of verb tense.

The recent-event preference is thought to reflect the stage of the CIA in which referential expressions are co-indexed to the visually available referents to ground them in the real world, which is argued to take precedence on predictive processing (Abashidze & Knoeferle, 2021). In addition, the robustness of such preference may be partially attributed to the verb-noun association established by the visual scenes. The latter explanation seems to be compatible with the view that a semantic (and probably visual) priming mechanism may be a driver of predictive behavior (and especially, predictive fixations, Huettig & McQueen, 2011; Kukona et al., 2011; Kuperberg et al., 2003).

However, predicting that the experimenter will soon be sugaring the pancakes implies both a referential and a thematic-role incongruence with the recently inspected scene, as the CIA's notion of reference includes the entities' relation to one another (Knoeferle et al., 2005). Thus, it possible that persisting fixations to previously inspected patients reflect (i) referential incongruence, on the one hand – i.e., the strawberries are "generically relevant" semantic entities, due to grounding of the visual referents – but also (ii) thematic-role incongruence, on the other hand – i.e., the strawberries may be considered specifically relevant as patients due to the thematic relations conveyed by the visual scenes.

In other words, the recent-event preference may be due to a preference to both attend to inspected/mentioned over uninspected/unmentioned referents (a "recent-referent" preference, so to say), and to attend to the most likely upcoming referents – if we assume that the thematic relations conveyed by the visual scenes are perceived as the most likely to occur again in the near future. For instance, the same thematic roles conveyed visually may be preferentially re-assigned to the same referents in utterances sharing the same referential entities as the depictions.

In addressing these considerations, I hypothesize a RECENT-ROLE PREFERENCE, i.e., a tendency to re-assign visually conveyed thematic roles to the same referents in temporally coordinated utterances, that this dissertation puts to the empirical test. By maintaining the semantic entities constant and available in both spoken and visual stimuli, differences in

referential grounding between conditions are minimized to directly measure the role of visually encoded relations on (predictive) comprehension of thematic relations.

Evidence for a recent-role preference would indicate that the previously or concurrently inspected thematic-role representations increase the perceived likelihood of the same referent taking on the same role in present and future utterances. In line with this hypothesis, previous evidence has shown that the prior inspection of given thematic-role relations triggered effects of incongruence on-line both at the behavioral and neurophysiological level, suggesting that visually encoded relations influenced thematic-role representations of unambiguous sentences – all referential entities being equal (Knoeferle et al., 2014; Knoeferle & Crocker, 2005). However, this research did not assess comprehension of thematic relations post-sentence.

Crucially, the recent-event preference would fall short in explaining findings from the current work, as the differences in the referential grounding of distinct semantic entities are minimized, and the effects of visual-scene cues may only be attributable to differences in the "relational grounding" of the available entities. In contrast, the recent-role preference might have some explanatory power for the patterns observed in recent-event studies.

1.4   Cross-speaker variability in cue weighting

1.4.1   First- and second-language speakers

While the ability to decode and use syntactic information to derive meaning is relatively automatized in first-language (L1) and early second-language (L2) learners, later L2 acquisition and lower proficiency have been found to be associated with lower ability to perform automatized parsing of morphological and syntactic cues (Díaz et al., 2016; Hopp, 2015b; Ojima et al., 2005; Pakulak & Neville, 2011; Tanner et al., 2013; Veríssimo et al., 2018).

The SHALLOW STRUCTURE HYPOTHESIS (SSH) suggested that L1-L2 processing differences may be attributed to varying ability to decode and use syntactic integration during comprehension (Clahsen & Felser, 2006, 2018). This dissertation refers to the ability to decode and make use of cues from different sources as "sensitivity". In particular, L2 (vs. L1) speakers may be less sensitive to grammatical information, while the ability to make use of non-syntactic cues may be more similar across both groups. The SSH is compatible with the view that sentence processing may rely on combinatorial and/or inference-based strategies, as argued by dual-route accounts for L1 speakers. Further, the SSH argues that sensitivity to specific cues may vary systematically across L1 and L2 speakers.

For example, Hopp (2015b) examined the integration of case and semantic cues in driving thematic-role prediction in L2. In a VWP design similar to Kamide et al. (2003), L1 and L2 speakers of German listened to thematically unambiguous spoken SO/OS sentences (e.g., SO: *Der Wolf tötet gleich den Hirsch* 'The wolf will soon kill the dear'). L1 speakers were able to

anticipate upcoming referents based on the integration of case and semantic cues regardless of word order, contrary to Kamide et al.'s (2003) results. Instead, L2 speakers integrated semantic, but not case cues from the first NP and the verb, as they always anticipated the second NP to take on the patient role. This evidence points to lower sensitivity to case cues for prediction during sentence processing in L2, resulting in an overwhelming reliance on word-order cues (but cf. Schlenter & Felser, 2021).

Further, individual differences have been shown to play a pivotal role in determining what information is relied on during comprehension. This evidence supports the view that sensitivity to specific cues may go beyond categorical L1-L2 distinctions in explaining differences in thematic-role comprehension and prediction across speakers. While the role of skills such as memory, interference control, and reading span have been examined more systematically (Cunnings, 2017; Hopp, 2014; Lewis & Vasishht, 2005; Linck et al., 2014; Tse & Altarriba, 2014; Yadav et al., 2022; Zhou et al., 2017), little research so far has looked at sensitivity to specific linguistic cues (Hopp, 2015a; Kilborn, 1992).

For instance, Hopp (2015a) explored whether syntactic- and semantic-integration abilities predicted reliance on syntactic and semantic cues in L2 sentence processing. A word-monitoring task was administered to assess grammatical- and semantic-integration skills (Marslen-Wilson & Tyler, 1980). Eye movements were monitored during reading of garden-path sentences in which different cues supported or opposed the initial misinterpretation. L2 speakers showed increased difficulty in the "pronoun" condition (e.g., *When the girl was praying he made some funny noises*) relative to the control condition (e.g., *When the girl was praying, the boy made some funny noises*), indicating that the pronoun was initially misinterpreted as being the direct object, though it was unambiguously marked as a subject. In addition, syntactic-integration abilities were associated positively with the use of case information, and negatively with the use of plausibility cues. Moreover, syntactic- and semantic-integration abilities correlated negatively with one another. These results suggest a potential trade-off in the reliance on syntactic and non-syntactic information for comprehension across speakers, which may be determined by varying sensitivity to such cues at the individual level.

To summarize, varying reliance on different cue types may underlie L1-L2 differences in comprehension, and sensitivity to each cue type may predict to what extent such cues are relied on in L1 and L2 comprehension alike. It remains to be determined whether L2 speakers' reliance on inference-based over combinatorial processing strategies for comprehension extends to the use of visual information, and how the latter relates to the use of word-order and morphosyntactic cues for the comprehension of thematic relations.

1.4.2   The adult lifespan

Changes in language processing across the adult lifespan have long been attested. While a consensus has been reached that some linguistic abilities decline with age (e.g., lexical retrieval, Verhaegen & Poncelet, 2013; predictive processing, Federmeier et al., 2010), research on syntactic processing has yielded more mixed evidence.

On the one hand, age-related declines in the sensitivity to syntactic cues during comprehension have been reported (e.g., Beese, Vassileiou, et al., 2019; Christianson et al., 2006; Poulisse et al., 2019). For example, Stine-Morrow et al. (2000) tested younger and older adults' word-by-word reading times and comprehension of subject- and object-extracted relative clauses in English (e.g., subject-extracted: *The pilot that admired the nurse dominated the conversation*). While younger adults allocated more time to read object- (vs. subject-)extracted relatives, older adults did not, and understood them less accurately. This suggests that older adults may have lower sensitivity to syntactic cues than younger adults. In turn, older adults' comprehension was more biased by the thematic relations conveyed by word order.

On the other hand, declines in syntactic comprehension across the lifespan have not been shown reliably. For example, Hardy et al. (2020) (Experiment 1) reported that younger and older participants showed comparable effects in a syntactic priming paradigm, indicating that older adults can maintain and generate previously activated syntactic representations.

In addition, older adults' crystallized knowledge becomes ever richer and may be exploited strategically to support comprehension (Stine-Morrow, 2007). This may be because inference-based processing is more available and/or computationally less demanding for older adults (Amichetti et al., 2016; Beese, Werkle-Bergner, et al., 2019; Luo et al., 2007; Stine-Morrow et al., 1996). This evidence points to a "syntax-to-semantics" shift in the information sources used during comprehension. This view aligns with the rationale of dual-route accounts by positing an age-related re-orienting of reliance on either combinatorial or inference-based processing routes.

For example, DeDe (2015) tested older and younger participants' sensitivity to syntax and animacy cues by measuring self-paced reading times and agent identification of subject- and object-extracted relative clauses with an (incongruent) inanimate subject head noun, or an (incongruent) animate direct-object head noun. Incongruent animacy cues were more disruptive for older than younger adults' comprehension, which was interpreted as "older adults [experiencing] greater processing disruptions when linguistic input differs from their knowledge-based expectations of language" (DeDe, 2015, p. 495). Further, Beese, Werkle-Bergner, et al. (2019) examined to what extent younger and older adults benefited from syntactic and semantic relations during sentence processing. The benefit from syntactic constraints was larger in younger than in older adults, suggesting that syntactic information

may be relied on less with increasing age. In contrast, the advantage from semantic constraints was comparable at young and old age.

Radvansky (1998) and Zwaan & Radvansky (1998) explained such "syntax-to-semantics" shift in terms of ability to retrieve information from SITUATION MODELS, i.e., from "the representations of situations as they would exist in a real or possible world. As such, [situation models] include […] inferences generated from world knowledge" (p. 1). Crucially, older adults have been shown equal or better memory performance for information extracted from situation models relative to younger adults (Radvansky et al., 2001, 2003). In particular, some research has investigated age-related changes in the reliance on perceptual cues in building situation models.

For instance, Dijkstra et al. (2004) presented younger and older adults with written sentences followed by object depictions and asked them to judge whether the depicted object had been mentioned in the sentence. The depiction represented the object (e.g., a bird) either with the same or different shape features as those implied by the sentence (e.g., in flight vs. perched). Older (vs. younger) adults exhibited stronger negative effects of shape incongruence on response times, indicating that they might build and rely on situation models to a greater extent. Madden & Dijkstra (2010) used a similar design, in which participants listened to sentences describing objects, then saw object depictions either congruent or incongruent with the shape features implied by the sentence and performed a picture-naming task. Older adults experienced greater benefits from congruent (vs. incongruent) shape features than did younger adults. Together, this evidence indicates that the ability to build mental situation models and exploit them to support comprehension is stable, and may even improve, throughout the adult lifespan.

To summarize, the evidence is mixed as to what extent increasing age influences the comprehension of syntactic structures, especially those presenting noncanonically ordered constituents. Comprehension seems to be supported by different information sources throughout the adult lifespan. Particularly, while sensitivity to syntactic cues may (or may not) decline with increasing age, the ability to integrate semantic and perceptual features to shape mental representations is fully maintained, and it might be used strategically by older adults to support comprehension.

1.5   Aims, Research Questions, Hypotheses

The current work investigated the relative contribution of cues from morphological case, linear word order, and visual scenes to the comprehension of thematic relations in unambiguous spoken relative clauses in German. Case cues on the relative pronoun assigned to it either the subject- or direct-object grammatical role, while word order was kept constant across conditions. This resulted in canonically and noncanonically ordered relative-clause structures (see 1a-b, repeated below). The spoken sentences were presented either in isolation,

or alongside a visual depiction conveying either the same or the opposite thematic-role representations as the sentence. We measured the relative contribution of each cue to such representations with off-line and on-line agent-identification tasks.

1	a	*Das ist der Koch,*	***der***	*die Braut*	*verfolgt.*
		This is the.NOM cook	who.NOM	the.ACC bride	follows
		This is the cook who is following the bride.			

	b	*Das ist der Koch,*	***den***	*die Braut*	*verfolgt.*
		This is the.NOM cook	whom.ACC	the.NOM bride	follows
		This is the cook whom the bride is following.			

The overarching question addressed by this thesis was: How do visual scenes influence thematic-role representations of canonically and noncanonically ordered clauses?

Based on the CIA, we assume that temporally coordinated visual and linguistic entities are co-indexed and that the thematic relations between referents conveyed by each of them are integrated. Thus, we expect visually encoded relations to influence the comprehension of the thematic relations of temporally coordinated utterances (Publications 1, 2, 3).

If visual effects during comprehension are robust, we expect the previously attested contribution of visual-scene cues to thematic-role assignment to replicate in the current setting, which was designed to provide potentially infertile ground for such effects to emerge. For one, the visual contribution was assessed for the comprehension of unambiguous structures. This is because the syntax suffices to derive a fully specified sentence representation, making the visual input technically superfluous for thematic-role assignment. Further, visual effects were measured from a purely linguistic task unrelated to the visual component, aiming to address potential concerns as to the task- and stimulus-relatedness of visual effects in previous work. Lastly, the effects of congruent and incongruent visual scenes on agent identification were assessed against a neutral baseline presented with no visual scenes, to tease apart the respective contribution of both congruent and incongruent scenes to thematic-role representations.

As discussed, such contribution may be explained by a recent-role preference, i.e., a tendency to re-assign visually conveyed thematic roles to the same referents in temporally coordinated utterances. Evidence for such preference would indicate that the previously or concurrently inspected thematic relations may increase the perceived likelihood of the same referent taking on the same role in present and future utterances.

When borne out (i.e., when visually and linguistically encoded relations converge), the recent-role preference may facilitate processing and improve recall of the thematic roles of the

sentence. According to the CIA, congruent visual and linguistic thematic relations should be seamlessly co-indexed and integrated, resulting in earlier thematic-role prediction and in faster and more accurate agent identification.

When disconfirmed, the recent-role preference may delay thematic-role prediction, increase response times, and possibly decrease accuracy of agent identification. According to the CIA, detecting thematic-role incongruence should trigger an attempt to reconcile visual and linguistic input to build a coherent thematic-role representation. Measuring to what extent conflicting visual thematic relations affect such representations post-sentence and on-line enables us to further specify the outcomes and dynamics of such reconciliation process. For example, observing effects of incongruent visual scenes on thematic-role assignment of canonical sentences would reveal that the recent-role preference may override the subject advantage in determining the likelihood of a referent taking on a specific thematic role in a temporally coordinated utterance.

Furthermore, the recent-role preference may interact with the subject advantage. For example, the difficulty associated with thematic-role assignment of noncanonical structures may be neutralized by congruent visual scenes, as the predictability of the first mentioned character taking on the patient role may be increased by the recent-role preference. Moreover, such difficulty may be exacerbated by incongruent visual scenes, which may add to the expectations engendered by the subject advantage as to the sentence's (erroneous) thematic roles. However, such interactions have not been shown consistently by previous work (e.g., Hopp, 2015b; Kamide et al., 2003).

In addition, this dissertation extended the applicability of predictions from the CIA to second-language speakers and speakers from across the adult lifespan by addressing the following questions:

1. How do visual scenes influence thematic-role representations of canonical and noncanonical sentences in L1 and L2 comprehension?

I hypothesized that sensitivity to case cues may predict reliance on visual cues in both L1 and L2 comprehension. Furthermore, I explored whether case sensitivity may predict agent-identification success beyond L1-L2 grouping, which would suggest that L1-L2 differences in comprehension strategies can be understood in terms of individual variability in the sensitivity to specific cues to thematic-role assignment (Publication 2).

2. How do visual scenes influence thematic-role representations of canonical and noncanonical sentences across the adult lifespan?

Based on the evidence for a shift towards inference-based over algorithmic processing with increasing age, I explored whether visual cues might be considered as one of the information sources that are relied on increasingly to build thematic-role representations with

increasing age. If so, congruent visual cues may improve agent-identification success and incongruent visual cues may hinder it to a greater extent with increasing age (Publication 3).

2   Overview of the Publications

The experimental work of the dissertation is composed of three manuscripts, which are currently published or undergoing peer-review in international academic journals. The current chapter provides an overview of the specific goals and outcomes of the empirical work underlying the thesis.

The empirical work of this dissertation presented subject-extracted and object-extracted relative clauses in German (see 1a-b). The spoken sentences were presented either in isolation, or alongside a visual depiction conveying either the same or the opposite thematic-role representations as the sentence. The design choices that varied across publications are specified in the respective summaries.

2.1   Publication 1

Single author; under review by *Journal of Psycholinguistic Research*.

The temporal dynamics of incongruence resolution for situated thematic-role prediction

Author: Carlotta Isabella Zona[1]

Author's affiliation:

(1)   Potsdam Research Institute for Multilingualism, University of Potsdam, Germany

Summary:

This study investigated the time course and the dynamics of the integration of conflicting case, word-order, and visual-scene cues during predictive thematic-role assignment. 69 speakers of German performed an agent-identification task in two steps. Firstly, participants inspected a visual scene and were asked to click on the depiction of the character performing a given action. Next, they listened to the spoken sentence and clicked on the portrait depiction (at the top corners of the screen) of the character performing the action in the sentence. The mouse-trajectory data were analyzed with two complementary non-parametric statistical approaches.

According to the recent-role preference, the thematic relations conveyed by the visual scenes should be perceived as the most likely to re-occur in the utterance, which should be observed immediately after sentence onset. As the main clause unfolds (i.e., *Das ist der Koch…* 'This is the cook…'), the first-mentioned (masculine) character is suggested as the most likely agent of the upcoming relative clause due to the subject advantage. The interaction of recent-

role preference and subject advantage would be indexed by effects of incongruence between depicted agent and first-mentioned referent prior to the disambiguating pronoun, which may result in delayed thematic-role predictions.

This hypothesis was supported, as the temporary incongruence of recent-role preference and subject advantage affected mouse movements as early as ~320ms after sentence onset. The signatures of thematic-role prediction were only reliable when recent-role preference and subject advantage converged. While sentence-picture incongruence did not seem to affect later trajectories or comprehension accuracy, responses were significantly slower in incongruent (vs. congruent) trials.

The results provide evidence for the "recent-role" preference and point to its interaction with the subject advantage in determining the time course of thematic-role prediction.

Personal contribution: This work was fully conducted by me.

2.2   Publication 2

First author; published in *Languages*.

Integrating morphosyntactic and visual cues in L1 and L2 comprehension

Authors: Carlotta Isabella Zona[1] and Claudia Felser[1]

Authors' affiliation:

(1)   Potsdam Research Institute for Multilingualism, University of Potsdam, Germany

doi: https://doi.org/10.3390/languages8020111

Summary:

This study examined L1-L2 and individual differences in sensitivity to visual-scene, word-order, and case cues for comprehension. 45 L1 and 39 L2 speakers of German listened to the spoken sentences and answered written agent-identification questions, as in 10a-b.

10	a	*Wird*	*die Braut*	*verfolgt?*
		Is	the.NOM bride	followed
		Is the bride being followed?		

	b	*Wird*	*der Koch*	*verfolgt?*
		Is	the.NOM cook	followed
		Is the cook being followed?		

The sentences were presented either in isolation (Experiment 1) or alongside thematically congruent or incongruent visual scenes (Experiment 2).

We hypothesized that visual scenes contribute to thematic-role representations of unambiguous utterances, and that sensitivity to case cues may predict reliance on visual information in both L1 and L2 comprehension.

The results revealed robust effects of visual cues on agent-identification accuracy in both groups and for both clause types. Additionally, higher syntactic sensitivity supported accurate comprehension particularly when object-relative clauses were presented with incongruent visual scenes. Notably, syntactic sensitivity emerged as a better predictor of accuracy rates than the categorical L1-L2 distinction.

This evidence extends the applicability of current models of situated sentence processing to L2 speakers and replicates previous findings in strictly conservative experimental settings. The observed differences may be attributed to the individual weighting of syntactic/non-syntactic information. This result emphasizes the importance of varying sensitivity to specific cue types in L1 and L2 sentence comprehension.

Personal contribution: I conceived the experimental set-up, designed, and programmed the experiment, and recruited the participants. I programmed and performed the statistical analyses of the data and interpreted the results. I wrote the first full draft of the manuscript and also contributed to editing and finalizing the manuscript.

Contribution of co-authors: Claudia Felser contributed to conceiving the experimental set-up, to interpreting the results, and to editing and finalizing the manuscript.

2.3   Publication 3

First author; under review by *Psychology and Aging*.

The weighting of syntactic versus visual-context information during comprehension: A lifespan perspective

Authors: Carlotta Isabella Zona[1] and Jana Reifegerste[1,2]

Authors' affiliation:

(1)   Potsdam Research Institute for Multilingualism, University of Potsdam, Germany

(2)   Department of Neuroscience, Georgetown University, USA

Summary:

This study investigated the lifespan trajectory of the use of visual-scene, word-order, and case cues for comprehension. 154 native speakers of German between 18 and 75 years of age listened to the spoken sentences and answered written agent-identification questions (see 10a-

b). The sentences were either presented in isolation (baseline block) or alongside thematically congruent and incongruent visual scenes (visual-cue block).

We hypothesized that visual information influences thematic-role assignment of unambiguous relative clauses, and it may do so more in older than in younger adults' comprehension. We further explored whether the additional competition of word-order and case cues might result in lower agent-identification accuracy, and potentially in higher reliance on visual information, especially at old age.

Sentence-picture congruence led to higher accuracy and incongruence to lower accuracy compared to the no-scene baseline. The size of incongruence effects increased with age and was influenced by participants' interference control. However, differences in the comprehension accuracy of canonical versus noncanonical structures were smaller in older than in younger adults.

The results support the view that sentence comprehension difficulties at old age are triggered by the competition of visual-scene and case cues relatively more than by the competition of word-order and case cues. This suggests that increased reliance on perceptual cues, rather than lower sensitivity to syntactic cues, may be a driver of age-related changes in sentence comprehension.

Personal contribution: I conceived the experimental set-up, and I designed and programmed the experiment. I programmed and performed the statistical analyses of the data and interpreted the results. I wrote the first full draft of the manuscript and contributed to editing and finalizing the manuscript.

Contribution of co-authors: Jana Reifegerste recruited the participants, and contributed to programming parts of the experiment, to interpreting the results, and to editing and finalizing the manuscript.

3   Publication 1

Under review by *Journal of Psycholinguistic Research*.

The temporal dynamics of incongruence resolution for situated thematic-role prediction

Carlotta Isabella Zona[1]

(1)   Potsdam Research Institute for Multilingualism, University of Potsdam, Germany

3.1   Abstract

Integrating visual and linguistic cues enables comprehenders to predict upcoming referents during situated thematic-role assignment, as captured by the Coordinated Interplay Account (CIA, Knoeferle & Crocker, 2006, *Cog Sci*). However, visually grounding distinct referential entities has been found to override prediction ('recent-event' preference, Knoeferle et al., 2011, *Front Psychol*). This study aimed to further define the dynamics and outcomes of resolving thematic-role incongruence within the CIA, as such incongruence may contribute to recent-event patterns. Visual-scene, word-order, and case cues were manipulated to convey competing thematic relations while the referential entities were kept constant. 69 speakers of German first inspected a visual scene and clicked on the agent of a given action. Next, they listened to a subject- (SO) or object-extracted (OS) relative clause (e.g., *Das ist der Koch, der/den die Braut verfolgt* 'This is the cook who follows the bride/whom the bride follows') conveying the same or opposite thematic relations as the depiction. Participants indicated the agent in the sentence by clicking on its depiction at each top corner of the screen. We measured mouse trajectories every 25ms and accuracy and latency of agent identification. The trajectories headed towards recently inspected agents ~300ms after trial onset. Thematic-role prediction (SO vs. OS) was only reliable when visual and word-order cues conveyed congruent relations. Such predictions were revised ~500ms after unambiguous case cues were presented, regardless of sentence-picture congruence. We interpret this evidence in terms of a 'recent-role' preference, i.e., a tendency to re-assign recently inspected thematic roles to the same referents in coordinated utterances. Our findings contribute to specifying the CIA mechanisms underlying the resolution of incongruent representations, extending the generalizability of its claims across stimuli, tasks, and measured variables. Further, the recent-role preference may contribute to explaining the robustness of recent-event patterns.

3.2   Introduction

One important task in sentence comprehension is to understand thematic relations – i.e., the relationships between sentence referents, or 'who is doing what to whom'. The current paper examines the use of morphological case, word order, and visual scenes during the comprehension of such relations.

On the one hand, visual cues have been extensively used to investigate the timing of information processing within the visual-world eye-tracking paradigm (VWP), based on the assumption that cognitive processing drives visual-attention allocation in tight temporal correlation (Allopenna et al., 1998; Cooper, 1974). This research has revealed that comprehenders do not passively extract information from different linguistic cues as they perceive them. Rather, the currently available information is used to infer the most likely continuation of the input (Clark, 2013; Ferreira & Chantavarin, 2018). This ability, known as

prediction, is triggered by semantically and syntactically constraining contexts (e.g., Altmann & Kamide, 1999, 2009; Ito et al., 2018; Kamide et al., 2003; Lau et al., 2022; Van Berkum et al., 2005). On the other hand, visual cues have been found to influence sentence processing and prediction, for example by grounding referential expressions in the real world (Abashidze & Knoeferle, 2021; Knoeferle et al., 2011), and by contributing to the resolution of referential (Sedivy et al., 1999; Spivey et al., 2002; Tanenhaus et al., 1995) and thematic-role ambiguity (Knoeferle et al., 2005; Knoeferle & Crocker, 2006).

Accounting for the reciprocal influence of visual and linguistic information during 'situated' sentence processing, the *Coordinated Interplay Account* (CIA, Knoeferle & Crocker, 2006) was put forth as a theoretical model describing a two-directional mechanism underlying visuo-linguistic integration. First, referential expressions steer visual attention towards relevant entities for visual grounding. Next, the inspected entities are used to resolve referential and thematic-role ambiguity, and may cause the current interpretation to change. When incongruence is detected, representations are indexed as 'false'. However, the time course and dynamics of resolving thematic-role incongruence in visual and spoken components of the input are not fully understood.

In addition to the evidence supporting the integration of visual and syntactic cues for thematic-role prediction, a *'recent-event' preference* has been reported, by which participants are found to fixate preferentially towards recently inspected referents over potentially upcoming ones (Abashidze & Knoeferle, 2021; Knoeferle et al., 2011). Such preference is attributed to the visual grounding of referential expressions, which precedes predictive processing.

The current study examines to what extent thematic-role incongruence contributes to the robustness of the recent-event preference by hypothesizing a *'recent-role' preference* – i.e., a tendency to re-assign thematic roles conveyed visually to the same referents in temporally coordinated utterances. The recent-role preference is put to the empirical test by fundamentally attempting to replicate recent-event findings while maintaining all referential entities equal, to single out the contribution of thematic-role incongruence to the dynamics and outcomes of situated thematic-role prediction.

3.3   Background

3.3.1   *Prediction and the subject advantage*

Traditional approaches to sentence processing have defined the information (or 'cues') used to understand thematic relations, and the timing of the integration of cues from distinct sources. On the one hand, structural models have argued that syntactic information may take precedence over non-syntactic information in guiding comprehension (Federmeier, 2007; Frazier, 1987; Frazier & Fodor, 1978; Gibson, 1998). On the other hand, constraint-based

models have argued that all information is weighted probabilistically for comprehension as it becomes available (Spivey-Knowlton et al., 1993; Spivey-Knowlton & Sedivy, 1995; Taraban & McClelland, 1988). The current study assumes that all information is extrapolated incrementally (i.e., on a constituent-by-constituent basis) and used to determine the relative likelihood of specific thematic-role configurations between constituents.

An important constraint determining such likelihood is linear word order. For example, the *subject advantage* – i.e., the tendency to assign the agent, rather than the patient role, to the first-mentioned referent in a sentence – is typically evidenced by slower, less accurate, and more effortful processing of noncanonically ordered object-before-subject (OS) versus canonically ordered subject-before-object (SO) structures (e.g., Karimi & Ferreira, 2016; Lau & Tanaka, 2021).

However, linear word order is the most reliable cue to thematic-role assignment in English (MacWhinney et al., 1984), such that comprehenders might have learned to rely on this cue more than others. In contrast, German makes extensive use of morphological case to map sentence constituents onto the verb's thematic grid through markings on determiners and via suffixes. Feminine and neuter nouns present non-contrastive nominative-accusative markings (*die-die* and *das-das*, respectively), resulting in situations of thematic-role ambiguity. In such situations, speakers of German exhibit a strong subject advantage and typically interpret Noun-Verb-Noun sentences as subject-first (Bader & Meng, 1999; Schlesewsky et al., 2000; Schriefers et al., 1995).

Instead, masculine nouns present fully contrastive markings for nominative (*der*) and accusative case (*den*), which may be used to unambiguously assign thematic roles to constituents. Thus, comprehenders may be hypothesized to rely on word order less than in English, as more reliable cues to thematic-role assignment are available (Bader & Meng, 2018; Schlesewsky & Bornkessel, 2004). However, the subject advantage has been reported from agent-identification tasks using unambiguous German case markings (Bader & Meng, 2018, 2023; Meng & Bader, 2021). For instance, Meng & Bader (2021) reported decreased accuracy of agent-identification of OS (vs. SO) sentences in German which were thematically unambiguous. Notably, these patterns were present regardless of whether the thematic roles were semantically reversible (e.g., 'The father hugs the uncle'), biased (e.g., 'The chef ruined the roast'), or non-reversible (e.g., 'The chef cleaned the pan'). This suggests a pivotal role of linear word order for thematic-role assignment and shows that unambiguous case cues may not override the subject advantage in delayed agent-identification tasks.

The current paper conceptualizes syntactic preferences, such as the subject advantage, as constraints on prediction (Kuperberg & Jaeger, 2016). This is because the endorsement of a preference facilitates the processing of incoming input that aligns with the preferred interpretation(s). In contrast, the processing of competing information disconfirms a previously held belief as to the structure of the sentence. The interpretation, and the

expectations that such interpretation engendered, need to be revised, resulting in a processing cost.

For example, Kamide et al. (2003) presented spoken SO and OS sentences (e.g., SO: *Der Hase frisst gleich den Kohl* 'The_{NOM} hare_{NOM} eats soon the_{ACC} cabbage_{ACC}', i.e., THE HARE WILL SOON EAT THE CABBAGE; OS: *Den Hasen frisst gleich der Hase* 'The_{ACC} hare_{ACC} eats soon the_{NOM} fox', i.e., THE FOX WILL SOON EAT THE HARE) while monitoring fixations towards depictions of the sentence referents. In response to SO structures, fixations towards the upcoming second noun-phrase (NP) were observed already at the verb, indicating that comprehenders integrated syntactic and semantic information from the first NP and the verb for prediction. However, anticipatory fixations were not observed in response to OS sentences, pointing to the role of noncanonical word order in delaying thematic-role prediction.

Importantly, while conclusions about prediction in sentence processing are often drawn from VWP studies, visually displayed information has extensively been shown to influence and enable predictive processing (Chambers et al., 2002; Farmer et al., 2007; Knoeferle & Crocker, 2006; Kukona et al., 2022; Spivey et al., 2002; Tanenhaus et al., 1995; Zhang et al., 2021), as reviewed in the next section. Thus, the extensive use of the VWP may have engendered testing effects and led researchers to overestimate the role of prediction for sentence processing in isolation (Huettig et al., 2011; Huettig, 2015; Huettig & Mani, 2016; Huettig & McQueen, 2011). For example, Huettig et al. (2011) showed that shape competitors of target objects (e.g., target: paddle; competitor: flute) were inspected early on when the materials were presented in pictorial form, but only at a delayed stage when the objects were presented as written words. Rather, written presentation yielded early inspections of semantic competitors (e.g., target: paddle; competitor: sailboat), suggesting that distinct presentation modalities (e.g., written word vs. depictions) and visual environments (e.g., different competitor types) may affect what information is relied on during situated sentence processing.

The current study addresses concerns as to the task-relatedness of visual effects for predictive thematic-role assignment by measuring prediction from a linguistic task unrelated to the visual displays and in the absence of the visual component.

3.3.2   *The Coordinated Interplay Account*

Research on *situated sentence processing* considers language as naturally immersed in a visual environment and has typically aimed to define the time course of the integration of visual and linguistic input during comprehension. Visual scenes have been found to be integrated with syntactic and semantic cues incrementally and predictively to resolve thematic-role ambiguity and to anticipate upcoming referents, and the *Coordinated Interplay Account* (CIA) has been proposed as a theoretical model describing a cyclic mechanism to account for this evidence. First, attention allocation is steered towards the visual referents

which are identified as relevant by the utterance. Second, the visual information currently being attended to and the referential expressions are co-indexed and verified against one another. Crucially, the sentence interpretation may then be revised based on the visual input if incongruence is detected, to build an internally coherent mental representation.

For example, Knoeferle et al. (2005) presented spoken SO/OS structures presenting initial thematic-role ambiguity (e.g., OS: *Die Prinzessin malt gleich der Fechter* 'The_NOM/ACC princess paints soon the_NOM fencer', i.e., THE FENCER WILL SOON PAINT THE PRINCESS). Fixations were monitored towards a character engaging as agent or as patient in unique actions (e.g., a princess painting a fencer and being washed by a pirate). In response to SO structures, participants fixated on the to-be-mentioned NP upon hearing the verb, revealing that visual and lexico-semantic cues were integrated to assign thematic roles predictively. In contrast, the signatures of thematic-role prediction were delayed in response to OS structures. These results point to the role of visual-scene cues in resolving temporary thematic-role ambiguity and suggest that visually encoded thematic relations and the subject advantage may interact in determining the time course of thematic-role prediction.

However, because syntactic ambiguity allows for two or more readings of the same structure – at least up to a point – the biasing effects of non-syntactic cues may more easily be detected as compared to unambiguous structures, in which only one interpretation is licensed by the syntax. Thus, it is conceivable that the contribution of non-syntactic information may be more prominent for ambiguity resolution than for sentence processing at large.

For example, the time course of the integration of unambiguous case and visual-scene cues for thematic-role prediction was investigated by Knoeferle & Crocker (2006). In Experiment 2, fixations towards depictions of two referents (e.g., a detective and a wizard) performing actions onto a patient character (e.g., a pilot) were monitored. Participants listened to OS structures (e.g., *Den Piloten verzaubert gleich der Detektiv* 'The_ACC pilot_ACC jinxes soon the_NOM detective', i.e., THE DETECTIVE WILL SOON JINX THE PILOT). The action in the sentence (jinxing) was stereotypically associated with the character depicted as though performing an unrelated action. Similar to the current paper, such design choice allowed the authors to tease apart the contribution of stereotypical knowledge and visual depictions to thematic-role prediction by pitting both cues against one another. Participants used visual scenes (more than stereotypical knowledge) to direct their gaze towards the referents identified as relevant by the scene. These results suggest that noncanonical OS clauses were processed predictively, likely due to unambiguous case cues, which may have reduced reliance on linear word order. Further, this points to the strength of visual effects relative to other non-syntactic sources, such as stereotypical knowledge.

Knoeferle (2007) compared the time course of thematic-role prediction of thematically ambiguous versus unambiguous SO/OS sentences in situated comprehension (e.g., SO ambiguous: *Die Frau Orange tritt in diesem Moment den Sir Zwiebel* 'The_NOM/ACC Mrs. Orange

kicks currently the_ACC Sir Onion', i.e., MRS. ORANGE IS CURRENTLY KICKING SIR ONION; SO unambiguous: *Der Herr Orange tritt in diesem Moment den Sir Zwiebel* 'The_NOM Mr. Orange kicks currently the_ACC Sir Onion', i.e., MR. ORANGE IS CURRENTLY KICKING SIR ONION). The visual entities were attended to as soon as the utterance identified them as potentially upcoming referents (i.e., one region earlier for unambiguous than for ambiguous structures). In addition, participants were equally likely to fixate on the upcoming second NP regardless of sentence structure in response to thematically unambiguous (but not ambiguous) OS sentences.

Together, these results suggest that unambiguous case cues may override the subject advantage during situated thematic-role prediction.

3.3.2.1   *Thematic-role incongruence in sentence-picture integration*

Some research has examined the outcomes and time course of the integration of incongruent thematic relations conveyed by visual versus linguistic components of the input.

For instance, Knoeferle & Crocker (2005) used a self-paced reading task to show that the congruence of visually and linguistically encoded thematic relations is verified incrementally regardless of task requirements. Similar to the current approach, the authors measured visual effects on comprehension with a linguistic task that does not hinge on the visual component, and as such, it may be considered a less spurious and more conservative assessment of visual effects, as argued above. Written SO/OS sentences described previously inspected depictions either with congruent or incongruent thematic relations (e.g., SO: *Die Oma filmt soeben den Handelskaufmann nach dem Vertragsabschluss* 'The_NOM/ACC granny films currently the_ACC businessman after signing the contract', i.e., THE GRANNY IS CURRENTLY FILMING THE BUSINESSMAN AFTER SIGNING THE CONTRACT). As a baseline, the sentences were presented following depictions of both referents in a neutral position, causing sentence-picture incongruence in terms of the action. Thematic-role (in)congruence was only evident at the determiner of the second NP through unambiguous masculine nominative-accusative case markings, while action incongruence was evident at the verb. Whole-sentence reading times did not show to be affected by the visual component but increased selectively at the verb in both incongruent (vs. congruent) conditions. Thematic-role (vs. action) incongruence yielded subtler effects overall, which were evident within the verb and the adverb regions, but not at the second NP. This seems surprising because the region of the second NP signaled thematic-role incongruence. While the results suggest that the inspection of given thematic relations between referents may increase the perceived probability that such referents would take on the same roles in coordinated sentences, the time course of the integration of thematically incongruent components requires further investigation, as does the nature of the thematic-role representations resulting from such integration.

Directly addressing the resolution of incongruent input within the CIA, Knoeferle et al. (2014) compared the neurophysiological correlates of the integration of thematic-role and

action-verb incongruence during sentence-picture verification. Participants read SO sentences (e.g., *The journalist punches the gymnast*) after inspecting depictions which were either congruent (journalist punching gymnast) or incongruent with the sentence. Incongruence was realized either in terms of thematic relations (gymnast punching journalist), action verb (journalist praising gymnast), or both (gymnast praising journalist). The signatures of thematic-incongruence processing emerged 100 to 400 ms after the first NP, continued briefly (~100ms) into the verb region, and were later evidenced by broadly distribute negativities. Notably, such neurophysiological signatures were indistinguishable from those typically associated with the processing of low cloze probability words during reading in isolation.

These patterns suggest that recently inspected thematic relations may be construed as the most likely to occur again in coordinated sentences sharing semantic overlap with the inspected scenes. The CIA was updated in light of this evidence by including a truth-value index appended to each representation. For example, the representation of the gymnast-as-patient engenders a mismatch when it is co-indexed to *The gymnast…* (as first mentioned character) and therefore it is indexed as 'false'. Such representations are thought to potentially remain active, though 'the extent and content of these [representations] is unclear' (Knoeferle et al., 2014, p. 146). The current study examines whether, when, and how such 'false' representations influence thematic-role representations on-line and post-sentence.

3.3.2.2   *Recent-event and recent-role preferences*

Some evidence has been reported for a 'recent-event' preference – i.e., a tendency to fixate on recently inspected over potentially upcoming referents during situated sentence comprehension (e.g., Abashidze et al., 2014; Knoeferle et al., 2011).

For instance, Knoeferle et al. (2011) tracked participants' eye movements as they listened to locally ambiguous spoken sentences which either referred to a recently viewed scene (e.g., *Der Versuchsleiter zuckerte kürzlich die Erbeeren* 'The experimenter recently sugared the strawberries') or to a potential future action (e.g., *Der Versuchsleiter zuckert demnächst die Pfannkuchen* 'The experimenter will soon sugar the pancakes'). Participants were found to inspect the recent-event object (the strawberries) more than the potentially upcoming referent (the pancakes) regardless of verb tense.

The recent-event preference is thought to reflect referential grounding, or the stage of the CIA in which visual entities and referential expressions are co-indexed, which is likely to take precedence on the attempt to predict upcoming items. The robustness of such preference has also been partially attributed to the verb-noun association established by the visual scenes (Abashidze & Knoeferle, 2021). The latter explanation seems likely in light of evidence supporting the view that a semantic (and probably visual) priming mechanism may be a main driver of anticipatory behavior (and especially, predictive fixations, Huettig & McQueen, 2011; Kamide & Kukona, 2018; Kukona et al., 2011; Kuperberg et al., 2003).

As an example, Kukona et al. (2011) monitored fixations towards characters stereotypically associated to the sentence's verb as agents or as patients during spoken comprehension (e.g., *Toby arrests the crook*; depictions: policeman, crook). At the verb, participants were equally likely to fixate on either referent – though the agent role had already been assigned – suggesting that 'local thematic priming can play a highly robust, and potentially larger role in anticipation than sentence-level syntactic constraints' (Kukona et al., 2011, p. 31).

In addition to strengthening the concerns expressed above as to the role of visual priming for predictive sentence processing, these observations may contribute to explaining the robustness of the recent-event preference (Abashidze et al., 2019; Abashidze & Knoeferle, 2021). In recent-event studies, the prediction that 'the experimenter will sugar the pancakes' is both referentially and thematically incongruent with the recently inspected scene. Thus, it possible that persisting fixations to previously inspected patients reflect (i) referential incongruence, on the one hand – i.e., the strawberries are 'generically relevant' semantic entities, due to grounding of the visual referents – and (ii) thematic-role incongruence, on the other hand – i.e., the strawberries may be considered specifically relevant as patients due to the thematic relations presented by the visual scenes. This is because the CIA includes the entities' relation to one another in its notion of referential processing (Knoeferle et al., 2005, 2014).

In other words, the recent-event preference may be due both to a preference to attend to mentioned over unmentioned referents, and to attend to the most likely upcoming referents – assuming the thematic relations conveyed by the visual scenes are perceived as the most likely to occur again in the near future. The current paper addresses this possibility by hypothesizing and putting to the empirical test a *recent-role preference*, i.e., a tendency to re-assign visually conveyed thematic relations to the same referents in temporally coordinated utterances. By maintaining the semantic entities constant and available in both spoken and visual stimuli, differences in referential grounding between conditions are minimized to single out the contribution of visually encoded thematic relations to (predictive) thematic-role comprehension.

3.4 Methods

3.4.1 Design

The current study used spoken sentences in German containing a main clause either introducing canonical subject-extracted (OS) or noncanonical object-extracted (SO) relative clauses (e.g., SO: *Das ist der Koch, der die Braut verfolgt* 'This is theNOM cook whoNOM the bride follows', i.e., THIS IS THE COOK WHO IS FOLLOWING THE BRIDE; OS: *Das ist der Koch, den die Braut verfolgt* 'This is theNOM cook whomACC the bride follows', i.e., THIS IS THE COOK WHOM THE BRIDE

IS FOLLOWING). Prior to each sentence, participants inspected visual scenes depicting either the same (congruent) or opposite (incongruent) thematic relations as the sentence, and were asked to click on the character performing the action in the depiction (Step 1). As the sentence was being presented, participants moved their mouse towards two portrait-like depictions of the characters at each top corner of the screen to identify the character performing the action in the sentence (Step 2).

The dependent variables included x- and y-coordinates of mouse trajectories sampled every 25 milliseconds (ms) during Step 2, as well as total latency and accuracy of responses.

3.4.2   Participants

The participants were recruited through the online participant pool of the University of Potsdam, provided their informed consent to participate in the study and were compensated with 8€ for participating. The sample included 69 German native speakers between 19 and 39 years of age. Participants declared to have been exposed to German only until age 5, reported no history of psychiatric or neurological disorders, and normal (or corrected-to-normal) vision and hearing (including color vision). In addition, participants provided information as to their education, handedness, and additional spoken languages. For each additional language, participants provided information as to age and context of acquisition. All participants who spoke any additional languages spoke English (n = 65), which they had started learning in a classroom setting at 5-6 years of age. Four of 69 participants spoke no languages other than German. 24 of 65 participants who spoke English spoke no additional languages other than German and English. After English, the most spoken languages in our sample were French (n = 29) and Spanish (n = 9). An overview of the participants' information is provided in Table 1.

Table 1: Participant information.

Gender	n	Student	n	Education	n	Age	n
Female	49	Yes	58	7 to 15	32	19 to 20	7
Male	18	No	11	16 to 21	31	21 to 25	29
Other	2			21 to 26	6	26 to 30	23
				Mean (SD)	16.7 (3.5)	31 to 35	6
						36 to 39	4
						Mean (SD)	25.6 (4.7)

3.4.3   Materials

3.4.3.1   Spoken sentences

Participants listened to 24 target sentences beginning with a main clause introducing a male human character, followed by a masculine relative pronoun either in the nominative (i.e., subject role, SO condition) or accusative case (i.e., object role, OS condition). This was followed by a feminine-gendered noun referring to a female human referent, and finally by a transitive verb in the third person singular of the present tense (e.g., SO: *Das ist der Koch, der*

die Braut verfolgt 'This is the~NOM~ cook who~NOM~ the bride follows', i.e., THIS IS THE COOK WHO IS FOLLOWING THE BRIDE; OS: *Das ist der Koch, den die Braut verfolgt* 'This is the~NOM~ cook whom~ACC~ the bride follows', i.e., THIS IS THE COOK WHOM THE BRIDE IS FOLLOWING). The choice of including only masculine referents as pronoun antecedents was motivated by the goal of making the relative clause thematically unambiguous.

Pseudo-filler sentences (n = 16) presented the same structure as target sentences, but the pronoun antecedent was either a feminine, neuter or a plural noun, and the sentence was disambiguated either at the determiner of a masculine second referent by the nominative or accusative form (e.g., *Hier sind die Damen, die der Dieb erschießt* 'Here are the ladies whom the thief is shooting') or through subject-verb agreement (i.e., at the end of the sentence; e.g., *Dies sind die Leute, die die Dame photographiert*, 'These are the people whom the lady is photographing'). As for the target sentences, half of the pseudo-filler sentences included a subject- and half an object-extracted relative clause. Filler sentences (n = 32) were constituted by a main clause and either a temporal or causal subordinate clause, whose order relative to each other was counterbalanced across items.

3.4.3.2 *Visual scenes*

Visual depictions of scenes were photos of Playmobil© characters engaging in simple transitive actions against a white background. All pictures were cropped at a 1:1 width-to-height ratio and brightness and contrast were adjusted to make the depictions' visual salience as homogeneous as possible. For target (n = 24) and pseudo-filler (n = 16) trials, visual depictions always represented the two referents of the upcoming spoken sentence. Mismatching trials contrasted with the sentences only in terms of the thematic roles conveyed (i.e., the action direction). For filler trials (n = 32), the mismatch with the spoken sentence was realized in terms of number or position of referents or action-verb incongruence.

3.4.4 *Procedure*

The experiment was scripted in E-Prime (Psychology Software Tools Inc., 2016) and administered remotely with E-PrimeGO (Psychology Software Tools Inc., 2020). Participants received their Subject ID number over email and a link to download the experiment from the E-PrimeGO server, where the experimental data were automatically uploaded after the end of the procedure.

In a mouse-tracking agent-identification task in two steps, participants were presented with a fixation dot (500ms), a written transitive verb (500 ms), followed by a visual scene in which one character was carrying out the action specified by the verb onto another character (Step 1). Participants clicked on the character performing the action. In Step 2, participants listened to spoken sentences describing the visual scene either accurately or conveying the opposite action direction, resulting in congruent and incongruent conditions. Participants

clicked on one of two portrait depictions at each top corner of the screen to indicate the referent carrying out the action in the spoken sentence. The portrait depictions were cropped from the visual scenes to a medium-shot of each character's face at a 1:1 width-to-height ratio.

The outcome variables were total latency and accuracy of agent-identification responses, as well as x- and y-positions of the mouse cursor from Step 2 of each trial, which were collected every 25ms.

3.4.5   Statistical analyses

All statistical analyses were performed in R Studio (R Core Team, 2022). The data were imported into R with the *rprime* package (Mahr, 2020). Response times (RTs; natural-log) and accuracy rates were analysed with mixed-effects linear and logistic regressions of the binomial family performed with the *lme4* package (Bates et al., 2015). The analyses of response times and trajectories were limited to correctly answered trials. X-coordinates were operationalized as a proxy of competition between alternatives, and the analyses were limited to this outcome of interest to ensure statistical rigor. To prepare the data, the starting positions of all trajectories were re-aligned to start at the bottom center of the screen (x-position: 0) and end at the top left corner of the screen (x-position: -1). Thus, decreasing values for the x-position reflect attraction towards the correct response, and increasing values indicate attraction towards the incorrect response. Then, trajectories were temporally interpolated to represent each trajectory in an equal number (n = 101) of temporally equidistant steps ('time-normalized trajectories' henceforth). The data were plotted with *ggplot2* (Wickham, 2008).

The present study took a novel approach to analyzing mouse trajectories with two non-parametric approaches: cluster-based permutation testing and non-parametric bootstrapping. Cluster-based permutation was carried out with the *clusterperm* package (Barr, 2022), and non-parametric bootstrap analyses by adapting the statistical procedures outlined in Stone et al. (2021). For all comparisons, categorical factors were sum-coded to reflect the effect of OS (+0.5) versus SO (–0.5) clause structures, and of incongruent (+0.5) versus congruent (–0.5) sentence-picture pairs.

3.4.5.1   Cluster-based permutation testing

Cluster-based permutation is a multivariate logistic regression method which does not require the data to be aggregated, provides a statistically sound approach to incorporate time as a continuous variable, and can integrate physically motivated constraints of, e.g., hand movements (Barr, 2008; Maris & Oostenveld, 2007). While cluster-based permutation allows to reliably determine the presence or absence of an effect in the data, it does not enable us to establish the time course of temporal differences (Sassenhagen & Draschkow, 2019). The null hypothesis which can be rejected by this test is that the labeling of conditions is interchangeable. In other words, under the null hypothesis, we will observe roughly the same

distribution that we would have observed if we had randomly selected half of the trials from Condition A and half of the trials from Condition B, labelled them as Condition A, and labelled the remaining trials from both conditions as Condition B.

The statistical procedure was as follows: during the descriptive stage, we first computed the F statistics and p values for the effects of interest at each step using analysis-of-variance models. Second, we identified windows of contiguous significance ('clusters'), i.e., consecutive steps in which the effect of interest reached significance at a criterion of $p < .05$. Then, we computed and stored the 'cluster-mass statistics' as the sum of the p values detected in each cluster. In the inferential stage, we generated 1000 permutations of the dataset by randomly relabeling its conditions, each time providing an approximation of the distribution that would have been observed under the null hypothesis. From each the permutation, we computed the cluster-mass statistics (CMS, i.e., the sum of consecutive significant p values). Lastly, we compared the cluster-mass statistics obtained from the observed dataset to those obtained from the distribution obtained from the permutations.

3.4.5.2   Non-parametric bootstrapping

Non-parametric bootstrapping allows to estimate differences in effect onset times between two conditions or groups, as well as the uncertainty about these differences (Stone et al., 2021). Due to the nature of this test, we established a baseline condition for each comparison, set the values associated with its trajectories to zero by subtracting them from the trajectory values in the condition of interest, and then compared the condition of interest to zero. While the bootstrapping approach allows to estimate time-course differences between conditions, it assumes the presence of an effect (Stone et al., 2021).

Hence, both cluster-based permutation and non-parametric bootstrapping approaches are complementary because cluster-based permutation is first used to ascertain the presence (vs. absence) of an effect. In light of this information, the bootstrapping procedure can be applied to further define the temporal course of the detected differences.

3.5   Hypotheses

The current work puts the recent-role preference to the empirical test – i.e., a tendency to re-assign visually conveyed thematic relations to the same referents in temporally coordinated utterances. In doing so, this work aims to single out the contribution of visually encoded thematic relations to (predictive) thematic-role comprehension.

Evidence for a recent-role preference in the current setting would imply that the previously or concurrently inspected thematic-role representations increase the perceived likelihood of the same referent taking on the same role in present and future utterances – all referential entities being equal (Knoeferle et al., 2014). Crucially, the recent-event preference

would fall short in explaining findings from the current work, while the recent-role preference might have some explanatory power for the patterns reported by recent-event studies.

Further, we expect participants to exhibit a subject advantage, i.e., to overwhelmingly interpret the first mentioned referent as the agent, rather than the patient of an upcoming relative clause. In the off-line data, this should be indexed by slower and less accurate agent identification in OS (vs. SO) conditions, based on evidence for enhanced difficulty with processing and comprehension of OS (vs. SO) structures (Bader & Meng, 2018, 2023; Ferreira, 2003; Schlesewsky et al., 2000). Such misinterpretations have been argued to be due to increased memory-retrieval difficulties for OS (vs. SO) structures (Bader & Meng, 2018; Meng & Bader, 2021). In this scenario, SO-OS differences should be minimal in the present design, as the need for mnemonic operations was minimized by accepting responses as the sentence unfolded.

To support a predictive interpretation of the patterns yielded by the subject advantage, we should find all target trials to be associated with deflections towards the male referent from the masculine determiner (*Das ist der* 'This is the$_{NOM}$') – i.e., from ~790ms (SD = 181ms) after trial onset. Recall that the other referent was always female. However, only in SO trials the syntactic expectation is borne out at the relative pronoun (1500ms), while in OS trials, it is disconfirmed. Thus, if the subject advantage crucially contributes to thematic-role representations incrementally, we should observe trajectory differences between relative-clause types before 1500ms.

Prior knowledge of the subject advantage and the hypothesized recent-role preference within the CIA framework allow us to draw fine-grained predictions about the time course of conflict resolution (Table 2).

Table 2: Violation of recent-role preference and/or subject advantage in each condition at each step. The icons indicate whether the current step endorses ('V') or disconfirms ('X') the previously held interpretation preference.

Condition	Example Sentence	Visual Scene	Example Scene	*der* NP (790ms, SD = 181ms)	Rel Pron (1500ms)	Violation
Congruent SO ('who')	Das ist der Koch, der die Braut verfolgt / This is the cook who is following the bride	(SO)		✓	✓	none
Incongruent SO ('who')	Das ist der Koch, der die Braut verfolgt / This is the cook who is following the bride	(OS)		✗	✓	Recent-role preference
Congruent OS ('whom')	Das ist der Koch, den die Braut verfolgt / This is the cook whom the bride is following	(OS)		✗	✗	Recent-role and subject advantage
Incongruent OS ('whom')	Das ist der Koch, den die Braut verfolgt / This is the cook whom the bride is following	(SO)		✓	✗	Subject advantage

The recent-role preference is expected to yield mouse trajectories more skewed towards the recently inspected agent (vs. patient) from trial onset. For instance, Knoeferle et al. (2014) observed signatures of conflict as early as 100ms after trial onset. Further, incongruence of recently inspected and first-mentioned characters should yield conflict in the time window after the first (masculine) NP (see Table 2, column 3).

In congruent SO conditions, the prediction engendered by the recent-role preference converges with the subject advantage (column 3), and with the relative pronoun (column 4). In incongruent SO conditions, recent-role preference and subject advantage convey incongruent thematic relations – as the recently inspected patient is mentioned first. The subject advantage is later supported by the relative pronoun. In congruent OS conditions, the prediction engendered by the recent-role preference is at odds with the subject advantage, which in turn is disconfirmed at the relative pronoun. In incongruent OS conditions, the recent-role preference and the subject advantage support the same prediction, which is disconfirmed at the relative pronoun.

Thus, we expect to observe that congruent OS trials differ from congruent SO trials earlier than do incongruent OS trials. This is because both incongruent SO and OS conditions entail the violation of two distinct predictions, which are disconfirmed at different points in time (the recent-role prediction is disconfirmed by the subject advantage in SO structures, and the subejct advantage at the relative pronoun in OS structures).

The (ultimately misled) interpretation is maintained longer in incongruent OS than in congruent OS trials. This allows us to investigating the effects of a longer-held

misinterpretation (incongruent OS) versus of the disconfirmation of two consecutive predictions (congruent OS) on recovering from the violation of the subject advantage (see below). This may provide some insight as to the factors modulating to what degree incongruent component yield signatures of cognitive conflict.

Post-sentence, sentence-picture incongruence may yield longer response times, as well as lower accuracy of agent identification. Such patterns would indicate that representations indexed as 'false' upon encountering incongruence may linger to affect agent-identification outcomes (Knoeferle et al., 2014).

3.6   Results

Mean accuracy, response times, and x-positions over time are plotted in Figure 1. Descriptively, accuracy decreases in response to OS (vs. SO) clauses seemed steeper than the decreases in accuracy following incongruent (vs. congruent) scenes (Panel A). The effects of both factors seemed more homogeneous on response-time data (Panel B). Further, the visual inspection of mouse movements along the x-axis (Panel C) suggests that trajectories were largely determined by sentence-picture congruence in the initial stages after trial onset, while SO-OS differences seemed to become prominent at a later stage.
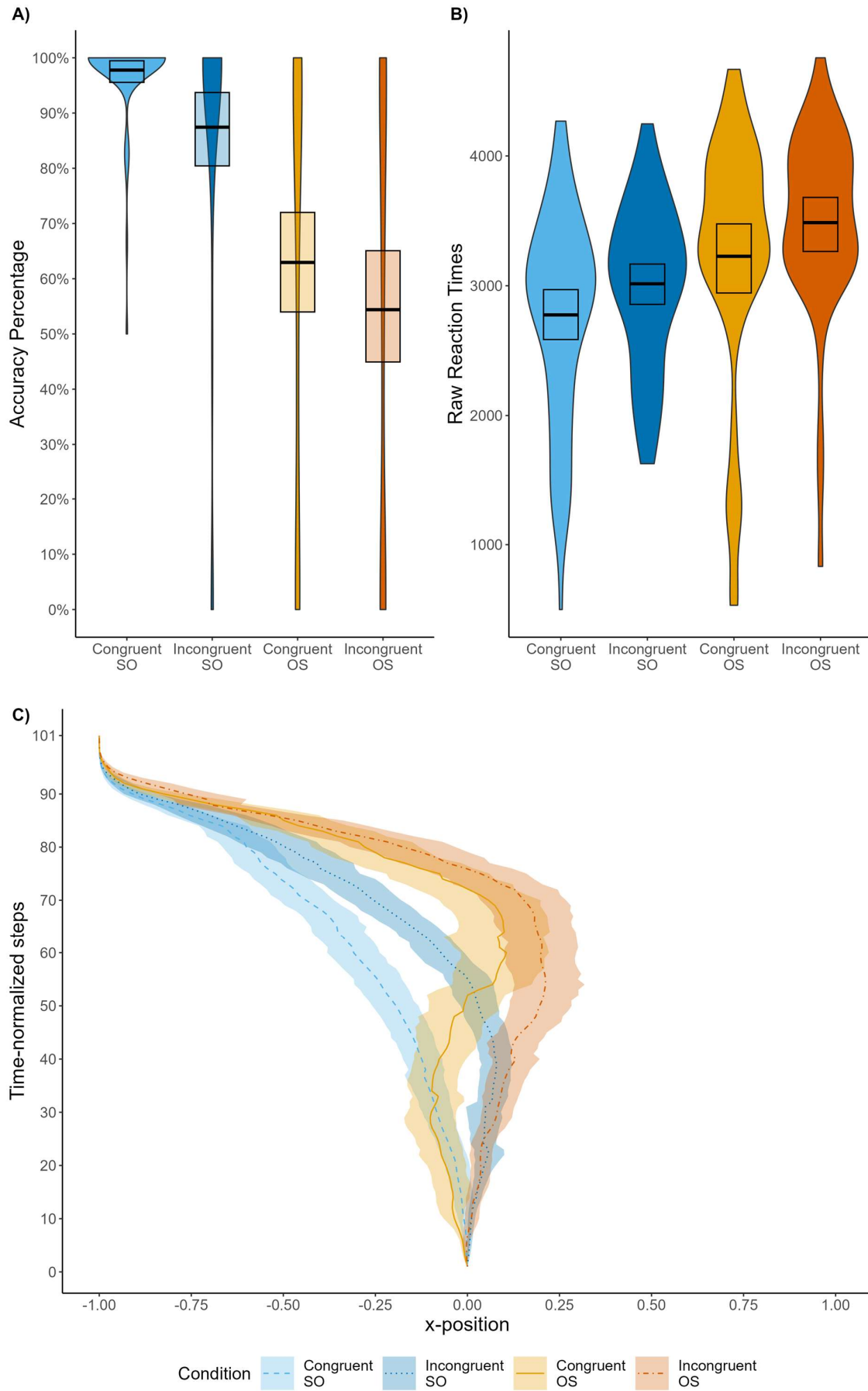
Fig 1 Descriptive results. Panel A: Average accuracy percentage by condition; Panel B: Average raw response times (in milliseconds) by condition; Panel C: Bootstrapped mean and 95% confidence limits of mouse positions on the x-axis over time by condition. Note that the trajectories were re-aligned to end in the top left corner of the screen

3.6.1   Off-line results

Both models fitted to accuracy and response-time data included Clause Type (SO, OS), Visual Scene (congruent, incongruent) and their interaction as fixed effects, as well as random by-participant intercepts and slopes for both fixed factors. The models' outputs are reported in Table 3. SO trials were answered significantly faster and more accurately than OS trials. This result is attributed to the subject advantage and broadly replicates well-attested increases in response times and lower accuracy of agent-identification of unambiguous noncanonical (vs. canonical) clause structures in German. Incongruent (vs. congruent) scenes yielded overall slower response times. Accuracy rates tended to be lower in response to incongruent (vs. congruent) scenes, though the effect did not reach significance. These patterns indicate that incongruent representations may have required more time to be integrated with one another and tended to result in less accurate thematic-role assignment relative to when both components of the input conveyed congruent relations. Broadly, this suggests that preceding Visual-Scene cues were integrated in the mental representation of thematic relations conveyed by the spoken sentences and affected the perceived likelihood of a referent taking on a specific thematic role in the coordinated utterance. Further, there was no evidence supporting an interaction of both factors on either dependent variable, suggesting that the visual scenes influenced responses regardless of the subject advantage.

Table 3: Regression coefficients of the logistic mixed-effects model (binomial family) fitted to accuracy rates (column 2) and of the linear mixed-effects model fitted to the natural-log transformation of the reaction times to correctly answered trials (column 3). For each predictor, estimates are reported in the first row outside brackets, standard errors are reported in round brackets, and *p*-values are reported in square brackets.

	Accuracy	Reaction Times
(Intercept)	3.696	7.922
	(0.715)	(0.045)
	[<0.001]	[<0.001]
Visual Scene	-0.992	0.039
	(0.631)	(0.012)
	[0.116]	[0.001]
Clause Type	-5.763	0.088
	(1.403)	(0.020)
	[<0.001]	[<0.001]
Visual Scene × Clause Type	0.495	-0.008
	(0.777)	(0.021)
	[0.524]	[0.693]
SD (Intercept Subject)	2.916	0.368
SD (Visual Scene Subject)	3.125	0.049
SD (SentCond1 Subject)	5.349	0.124
Cor (Intercept ~ Visual Scene Subject)	0.527	-0.692
Cor (Intercept ~ Clause Type Subject)	0.011	0.377
Cor (Visual Scene ~ Clause Type Subject)	-0.133	-0.505
SD (Observations)		0.169
Num. Obs.	1569	1179
R2 Marg.	0.287	0.013
R2 Cond.	0.890	0.829
AIC	829.2	-429.7
BIC	882.8	-373.9
ICC	0.8	0.8
RMSE	0.21	0.16

3.6.2   Cluster-based permutation results

The descriptive stage of the permutation test assessed the contiguous significance of the 101 time-normalized steps. For the Visual-Scene factor, a significant cluster extended between steps 30 and 82 from trial onset. For Clause Type, we observed a significant cluster between steps 41 and 91. The interaction of both factors was not significant in our dataset and was thus excluded from the inferential stage. The inferential analyses revealed the predicted effects as a function of both Visual-Scene (more skewed trajectories following incongruent than congruent scenes; CMS = 2841.08, p = .001), and Clause-Type factors (more skewed mouse trajectories in response to OS than SO clauses; CMS = 3025.28, p = .001).

3.6.3   Bootstrapped time-course differences

First, we tested whether participants exhibited signatures of predictive thematic-role assignment due to the subject advantage. With this aim, we compared mouse trajectories yielded by OS versus SO clause types – i.e., across both Visual-Scene levels. In OS trials, which disconfirm the subject advantage, we expected to observe deflections of mouse trajectories towards the incorrect alternative before the relative pronoun (1500ms). We measured the

earliest point at which the trajectories yielded by OS trials were significantly different from those yielded by SO trials (which were set to zero) for 10 consecutive 25ms steps.

OS clauses yielded reliable deflections towards the incorrect response starting at 1275ms (CI: 1250 – 1525ms). This pattern provides only weak evidence for a subject advantage in predictive thematic-role assignment, because the variance of the effect onset did not fall entirely before the time-locked onset of the relative pronoun (1500ms – Figure 2, Panel A).

However, it is likely that the observed delay in predictive thematic-role assignment is due to the disconfirmation of the recent-role preference at the determiner in incongruent SO trials. This incongruence might have triggered different response dynamics compared to congruent SO trials, which entailed no violation. To assess this, we examined the onset times of subject-advantage based thematic-role prediction in incongruent OS versus congruent SO conditions only – as in both of these the recent-role preference and subject advantage conveyed the same, congruent thematic relations up to the relative pronoun (cf. Table 2). The trajectories associated with congruent SO trials were used as baseline for this comparison. Significant differences were observed as early as 800ms (CI: 750ms – 1100ms), indicating that participants did engage in predictive processing, provided that recent-role and subject advantage conveyed congruent thematic relations (see Figure 2, Panel B).

Together, these patterns suggest that the temporary incongruence of the thematic relations conveyed by the visual scene and by the subject advantage contributed to delay thematic-role prediction as compared to the dataset as a whole. This indicates that participants were less likely (and/or slower) to anticipate the first-mentioned referent as the agent of the utterance if the same referent took on the patient role in a recent depiction.
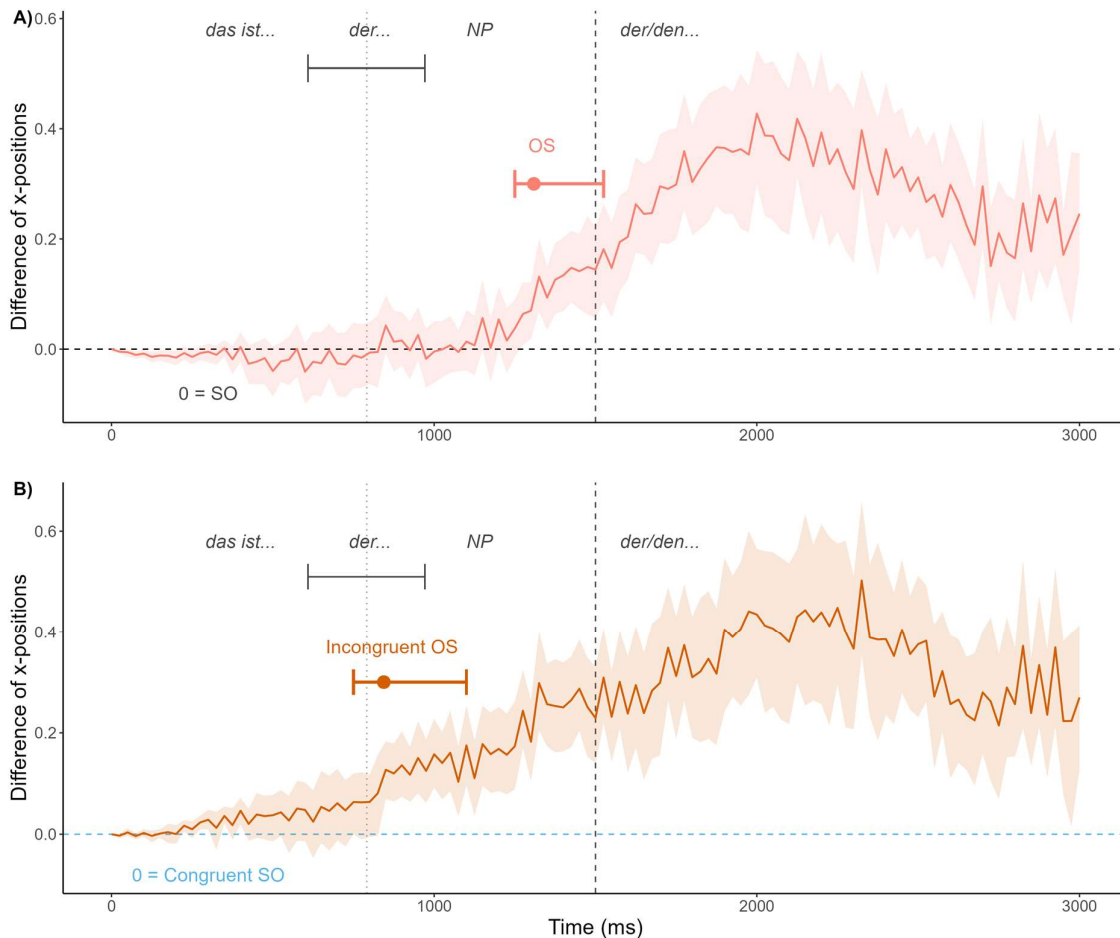
Fig 2 The time course of the effects of the subject advantage for thematic-role prediction. The vertical dotted lines represent the onset of the masculine determiner *der* 'the' (mean = 790ms, SD = 181ms), and the vertical dashed lines represent the onset of the relative pronoun (*der/den* 'who$_{NOM}$'/'whom$_{ACC}$'). The x-axis represents the trajectories in the condition used as baseline for each comparison. The plotted lines and shaded ribbons represent bootstrapped mean and 95% confidence limits of the difference in x-positions between the baseline condition and the condition of interest. The dots indicate the bootstrapped divergence point of each condition from the baseline. Error bars represent 95% bootstrapped confidence intervals of the onset of the effect. Panel A) represents the difference in x-position in response to OS (vs SO) clauses across Visual-Scene levels. Panel B) represents the difference in x-position in response to incongruent OS (vs congruent SO) trials

Next, we tested our hypotheses as to the time course of the integration of conflicting representations from recent-role preference, subject-advantage, and morphological case during thematic-role prediction. Congruent SO trials were used as baseline and set to zero. The x-positions of all OS trials were flipped such that differences from trajectories in SO trials reflect the earliest moment of correct categorization. For incongruent SO trials, in contrast, the differences reflect deflections towards the competitor.

The trajectories of incongruent SO and congruent OS trials started exhibiting deflections from the baseline 309ms (CI: 100ms – 425ms) and 316ms (CI: 150ms – 475ms) after trial onset, respectively, with no difference between conditions (Figure 3). The lack of differences in the

onset times of both effects is unsurprising because both conditions were identical up to the relative pronoun. Both conditions presented a temporary incongruence of the thematic relations conveyed by recent-role preference and subject advantage.

In contrast, the trajectories in incongruent OS trials started diverging from the baseline 1957ms (CI: 1750ms – 2175ms) after trial onset. The time course of these differences closely supports our hypotheses as to the interaction of the recent-role preference and the subject advantage on thematic-role prediction.
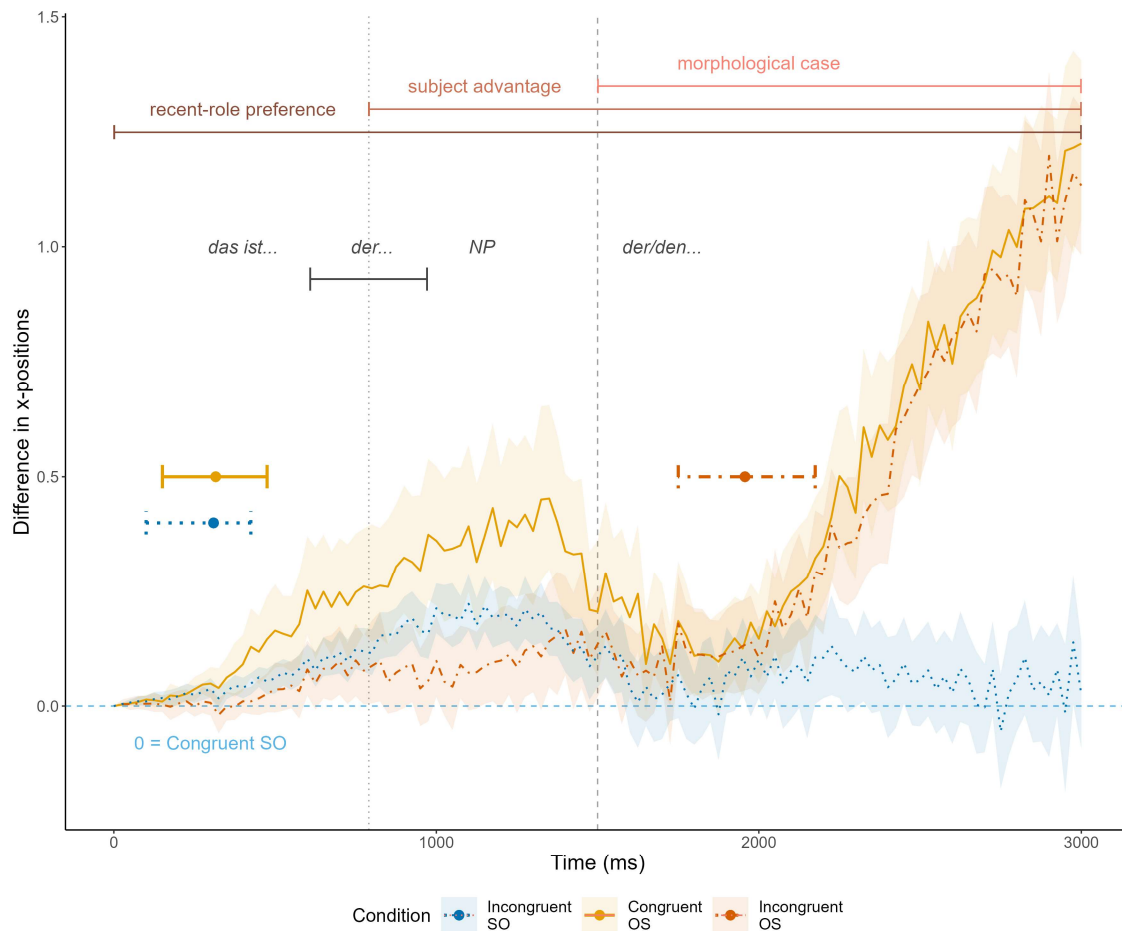
Fig 3 The time course of the effects of recent-role preference and subject advantage for thematic-role prediction. The vertical dotted line represents the onset of the masculine determiner *der* 'the' (mean = 790ms, SD = 181ms), and the vertical dashed line represents the onset of the relative pronoun (*der/den* 'who$_{NOM}$'/'whom$_{ACC}$'). The x-axis represents the baseline (congruent SO). The x-positions of OS trials were flipped, such that deflections from the baseline reflect the earliest moment of correct categorization. The error bars at the top represent our hypotheses as to the cues being integrated in each region. The plotted lines and shaded ribbons represent bootstrapped mean and 95% confidence limits of the difference in x-positions between the baseline condition and the condition of interest. The dots indicate the bootstrapped divergence point of each condition from the baseline. Error bars represent 95% bootstrapped confidence intervals of the onset of the effect

Lastly, a previous disconfirmation of the recent-role preference might influence later recovery from the disconfirmation of the subject advantage, for better or worse. On the one hand, violating two predictions consecutively may delay thematic-role assignment, as the interpretation needs to be revised twice. However, this implies that the recent-role preference is ultimately supported, and re-activating the discarded representation might facilitate recovery from the disconfirmation of the subject advantage. To assess this, the earliest moment of correct categorization was compared in congruent and incongruent OS trials. Incongruent (instead of congruent) SO trials were selected as baseline for a more conservative estimation. For this comparison, we examined only trajectory segments between 1500ms and

3000ms. The latter cutoff was necessary to avoid failures in the bootstrapping procedure which may arise with sparser data points collected at the end of the trial, when most participants have already responded (Stone et al., 2021).

Our results did not provide conclusive evidence as per the latter question, as both conditions showed largely overlapping patterns. The earliest time window of correct categorization was 2175ms for congruent OS (CI: 2050ms – 2250ms) and 2075ms (CI: 2100ms – 2350ms) for incongruent OS trials. The confidence intervals of the difference in divergence points between both conditions included zero (CI: -100ms – 175ms; Figure 4). We address possible reasons for this null result in the following section.
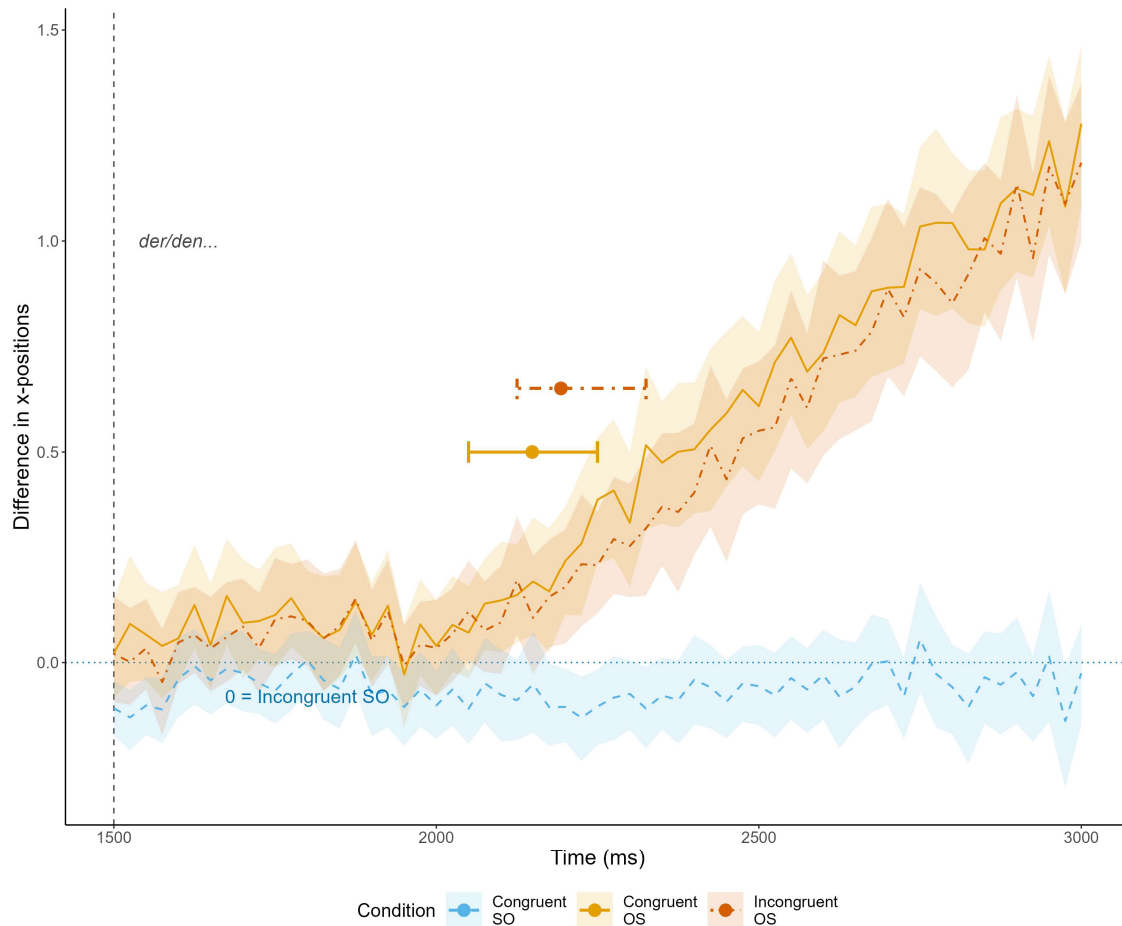
Fig 4 The time course of correct categorization in congruent and incongruent OS trials. The vertical dashed line represents the onset of the relative pronoun (*der/den* 'who$_{NOM}$'/'whom$_{ACC}$'). The x-axis represents the baseline (incongruent SO). The x-positions of OS trials were flipped, such that the effects reflect the earliest moment of correct categorization. The plotted lines and shaded ribbons represent bootstrapped mean and 95% confidence limits of the difference in x-positions between the baseline condition and the condition of interest. The dots indicate the bootstrapped divergence point of each condition from the baseline. Error bars represent 95% bootstrapped confidence intervals of the onset of the effect

3.7   Discussion

The present study investigated the time course and outcomes of the integration of competing cues from visual scenes, word order, and morphological case for thematic-role assignment of spoken relative clauses in German. In doing so, we hypothesized a recent-role preference – i.e., a tendency to re-assign recently inspected thematic roles to the same referents in temporally coordinated utterances.

The key findings can be summarized as follows: We found support the recent-role preference, as sentence-picture incongruence (vs. congruence) resulted in increased conflict between thematic-role representations during thematic-role assignment, resulting in longer

response times, yet only numerical decreases in response accuracy. Further, we measured a subject advantage, as SO (vs. OS) clauses yielded faster and more accurate responses, and word-order cues were used to assign thematic roles predictively. Importantly, when recent-role preference and subject advantage conveyed competing (vs. converging) thematic relations, thematic-role prediction was significantly delayed.

Next, the results are discussed in turn according to their temporal unfolding.

First, significant mouse-trajectory skews towards the response supporting the recent-role preference were measured as early as ~300ms after trial onset. This suggests that the recently inspected agent character was considered as the most likely agent of a coordinated sentence presenting unconstraining linguistic cues (e.g., "This is…").

Second, when the first (masculine) referent was mentioned, the agent role tended to be assigned to this referent. Importantly, however, the agent role was assigned predictively to the first referent only when the recent-role preference and subject advantage conveyed congruent relations. In contrast, the effects of the subject advantage on thematic-role prediction were not reliable across both Visual-Scene levels.

These patterns support the hypothesized recent-role preference and point to its interaction with the subject advantage in shaping thematic-role representations on-line. Visually encoded thematic relations modulated the time course and strength of the subject advantage – i.e., of the perceived likelihood that the first mentioned character will be the agent, rather than the patient of the action. In other words, inspecting a character as patient made participants less prone to assume that the same character would be the agent of the action, though it was mentioned first. The evidence supporting this conclusion seems particularly convincing because thematic-role prediction was found to be unreliable not only in conditions of incongruent recent-role preference and subject advantage, but in the dataset as a whole.

The CIA's notion of reference includes referents, actions, and the referents' thematic relation to one another established by the action. Further, the CIA predicts that thematic-role representations may be indexed as 'false' when an incongruence of visual and spoken components of the input is detected during the co-indexing stage (Knoeferle et al., 2014). However, the consequences of indexing representations as 'false' are not well understood. The present study extends the evidence for an incremental verification process underlying sentence-picture verification provided by Knoeferle et al. (2014) by showing on-line effects of competition between visually and linguistically conveyed thematic relations during a linguistic task unrelated to sentence-picture congruence. We further add to Knoeferle et al. (2014) by presenting stimuli in the spoken modality, and in German, which enabled us to model the competition of word-order and case cues for thematic-role comprehension. Thus, the present findings speak to the generalizability of a mechanism indexing incongruent representations as 'false' across tasks, languages, indices of cognitive competition (EEG

vs. mouse trajectories), and in lack of an explicit comparison to be performed (see also Knoeferle & Crocker, 2005).

According to the CIA, and based on evidence from recent-event studies, the semantic entities may be visually grounded before the attempt to predict upcoming referents. Based on the CIA's definition of reference, we hypothesized that recent-event findings may reflect both the visual grounding of semantic entities (a 'recent-referent' preference, so to say), and the grounding of the thematic relations between entities ('recent-role' preference). In other words, the recent-event preference may reflect both the tendency to attend to recently inspected (vs. non-inspected) referents, and to anticipate the most likely referents for specific roles, based on recently inspected relations. Crucially, the current study aimed to minimize the differences in the visual grounding of semantic entities presented by recent-event studies. This feature allowed us to single out the effects of visually grounding thematic relations between referents during thematic-role assignment. Thus, the effects of the recent-role preference in the current setting are interpreted in terms of the recently inspected thematic relations increasing the perceived likelihood that the same referents will take on the same thematic roles in current and future utterances showing overlapping semantic entities.

The current results may contribute to explaining the robustness of the recent-event preference (Abashidze et al., 2019; Abashidze & Knoeferle, 2021). The convergence of both tendencies – to attend to recently inspected (vs. non-inspected) referents, and to anticipate the most likely referents for specific roles – may have maintained recent-event fixations high regardless of the manipulations of frequency (Abashidze et al., 2019) and speaker-gaze cues (Abashidze & Knoeferle, 2021) designed to encourage reliance on syntactic (vs. visual-scene) cues for prediction. To provide conclusive evidence as to this possibility, future research may aim to directly compare the relative contribution of recent-role and recent-event preferences within the same design.

We return to the temporal unfolding of the observed effects. Next, the relative pronoun unambiguously assigning thematic roles was presented at 1500ms. In incongruent SO trials, the subject advantage was supported by the pronoun in nominative form and the trajectories proceeded relatively straight towards the correct response, with no statistical differences relative to congruent SO trials. Instead, in OS trials (congruent and incongruent), the subject advantage was disconfirmed by the relative pronoun in accusative form. Mouse trajectories started deflecting towards the correct response on average ~500ms after the onset of the pronoun. The time course of such shift was statistically indistinguishable in congruent and incongruent conditions.

The lack of (in)congruence effects on mouse trajectories after the first-mentioned referent is in line with the results from Knoeferle & Crocker (2005), who reported reading-time slowdowns limited to the sentence region signaling sentence-picture incongruence. In addition, word order may have been perceived as more important than the visual scenes by

our participants. This is because word order pertained to the spoken component, as did our task – though it led to the ultimately correct response only 50% of the times. Thus, it is possible that once a first-mentioned character was identified as the most likely agent, the influence of the previously inspected relations decayed steeply. After all, quickly discarding 'false' interpretations and beliefs is likely to be an important aspect of comprehension in our daily lives. This consideration may contribute to explaining the lack of later (in)congruence effects in the current task, contrary to the longevity of the visual effects reported by Knoeferle et al. (2014): Representations indexed as 'false' are likely to be more relevant to task performance throughout the trial in sentence-picture verification, as compared to agent-identification tasks. However, future research should aim to determine whether our null findings in later time windows might have been due to lack of statistical power in the present study.

Lastly, OS (vs. SO) clauses yielded slower and less accurate agent-identification responses post-sentence, as well as trajectories overall more skewed towards the incorrect response. These results replicate well-established findings for increased processing difficulty and decreased accuracy of agent identification of both ambiguous and unambiguous OS (vs. SO) clauses in German (Bader & Meng, 1999, 2018, 2023; Meng & Bader, 2021; Schlesewsky et al., 2000; Schlesewsky & Bornkessel, 2004). By enabling participants to respond as the sentence was being presented, the present setting minimized the confounds associated with delayed agent-identification in previous studies (e.g., Zona & Felser, 2023). This suggests that thematic-role representations derived from competing word-order and case cues may require more effort to be integrated and understood, rather than being more problematic to retrieve from memory for delayed tasks (Bader & Meng, 2018; Meng & Bader, 2021).

Sentence-picture incongruence (vs. congruence) yielded slower responses and only numerically decreased accuracy rates. Our results further suggest that the thematic-role representations indexed as 'false' may linger to affect post-sentence outcomes. In contrast to the current study, Knoeferle & Crocker (2005) reported no effects of sentence-picture congruence on total reading times of SO/OS clauses presenting initial thematic-role ambiguity. However, this discrepancy may be explained by differences in the design features and statistical comparisons performed. For example, Knoeferle & Crocker (2005) used written sentences and compared congruent and incongruent depictions to a baseline in which the characters were depicted in a neutral position. Instead, our results are in line with recent evidence showing significant effects of sentence-picture congruence on response accuracy and latency in delayed agent identification (Zona & Felser, 2023). The discrepancies in the results might be due to differences in the order with which the stimuli and tasks were presented in each study. While the visual scenes were presented before the sentence in the current study, Zona & Felser (2023) presented visual scenes and spoken sentences simultaneously, and administered a written binary agent-identification question in the following step. Because the time gap between the visual scene and the agent-identification task were comparable (one

step), the comparison of the results from both studies may point to simultaneous encoding of linguistically and visually conveyed relations as supporting the longevity of visual effects on thematic-role comprehension. However, the respective roles of recency and simultaneous encoding in determining the strength of visual effects on sentence comprehension still need to be addressed systematically.

The current study presents some limitations that motivate the need for further research. First, a replication of the current results should be provided in a more controlled laboratory setting, rather than online. For example, the participants' technological equipment could not be controlled for variables which have been found to affect mouse-tracking results, such as refresh rates and cursor speed (Schoemann et al., 2019, 2021). Further, to elucidate the role of sentence-picture congruence for predictive processing, a neutral baseline (presenting no visual scenes) may be included, as the effects of sentence-picture incongruence versus congruence may be due to facilitation from congruent information, interference from incongruent information, or both.

To summarize, our results show that comprehenders' interpretations are dynamically constrained by visual-scene, word-order, and case cues during predictive thematic-role assignment. Specifically, we reported a tendency to re-assign previously inspected thematic roles to the same referents in temporally coordinated utterances, which we dubbed 'recent-role preference'. Our results contribute to defining the underlying causes of the tendency to inspect recently inspected referents over potentially upcoming ones ('recent-event preference'), as well as to specifying the mechanisms underlying the integration of incongruent components of the input within the Coordinated Interplay Account (Knoeferle & Crocker, 2006).

4   Publication 2

The temporal dynamics of incongruence resolution for situated thematic-role prediction

Carlotta Isabella Zona[1] and Claudia Felser[1]

(1)   Potsdam Research Institute for Multilingualism, University of Potsdam, Germany

4.1   Abstract

This study investigates the relative weighting of morphosyntactic and visual cues in spoken-language comprehension, and whether this varies systematically within and between first (L1) and second language (L2) speakers of German. In two experiments, 45 L1 and 39 L2

speakers answered probe questions targeting the action direction of subject- and object-extracted relative clauses, which were presented either in isolation (Experiment 1) or alongside scene depictions either matching or mismatching the action direction expressed in the sentence (Experiment 2). We hypothesized that visual cues contribute to shaping meaning representations in sentence comprehension, and that sensitivity to morphosyntactic cues during interpretation may predict reliance on visual cues in both L1 and L2 comprehension. We found reliable effects of visual cues in both groups, and in response to both relative-clause types. Further, proxies of morphosyntactic sensitivity were associated with higher agent-identification accuracy, especially in response to object-extracted relative clauses presented with mismatching visual cues. Lastly, morphosyntactic sensitivity was a better predictor of accuracy rates than L1–L2 grouping in our dataset. The results extend the generalizability of models of visuo-linguistic integration across populations and experimental settings. Further, the observed sentence-comprehension differences can be explained in terms of individual cue-weighting patterns, and thus point to the crucial role of sensitivity to distinct cue types in accounting for thematic-role assignment success in L1 and L2 speakers alike.

4.2   Introduction

To understand sentences, listeners or readers need to process morphosyntactic cues that encode, among other things, sentence constituents' grammatical roles (e.g., subject, object), which helps in determining the direction of the action (i.e., 'who does what to whom', Van Gompel and Pickering, 2006). For instance, German makes extensive use of morphological case by marking pre-nominal determiners, adjectives, and nouns to indicate which grammatical function and possible thematic role (e.g., agent, patient) should be assigned to a noun phrase (NP).

Psycholinguistic research has shown that comprehenders may consider cues to interpretation from a variety of linguistic and non-linguistic information sources (Traxler 2011). Here, we use the term '(morpho)syntactic cue' to refer to interpretation cues that are encoded within the grammar. In contrast, 'non-syntactic' refers to all non-grammatical cues, which may be linguistic (e.g., semantic, pragmatic) or extra-linguistic in nature (e.g., visual context, world knowledge).

The relative weighting of different cue types, i.e., the extent to which comprehenders rely upon each type of cue to derive meaning, has been argued to differ between populations (e.g., Clahsen and Felser, 2006, 2018; Cunnings, 2017) and individuals (e.g., Yadav et al., 2022). Evidence from sentence comprehension in a second language (L2) suggests that non-grammatical information might be weighted more strongly in L2 than in the first language (L1). At the same time, morphosyntax might be relied on to a narrower extent, as captured by theoretical proposals such as the Shallow Structure Hypothesis (SSH, Clahsen and Felser, 2006, 2018) and Cunnings' Interference Hypothesis (Cunnings, 2017).

There is evidence that even L1 speakers may interpret language in a way that is merely 'good enough' for the current situation. That is, meaning representations may sometimes be computed that are incompatible with the linguistic input, especially if the linguistic stimuli represent marked structures (such as object extractions or passives), and if information from non-grammatical cues supports the misinterpretation (Ferreira, 2002). Evidence for the misinterpretation of noncanonical sentences has been interpreted as supporting dual-route models of sentence processing (Karimi and Ferreira, 2016), in which meaning can be derived either based on 'deep' grammatical parsing or on lexico-semantic and pragmatic information.

The core ideas of the dual-route approach to comprehension can be recast in terms of multiple-constraint models that allow for different weightings of information sources (see McRae and Matsuki, 2013 for a review), as has been argued by Clahsen and Felser (2018) for L2 processing. Conflicts between linguistic cues—where different types of co-occurring cues support different interpretations—have often been used to gauge the relative im-portance or weighting of each cue type for comprehension in L1 (Bates et al., 1982; MacWhinney and Pleh, 1997) and L2 speakers (MacWhinney, 2002, 2005). However, the interaction of linguistic and visual cues has not systematically been modeled in terms of competition in L1 or L2 processing. Moreover, the role of individual differences in cue weighting for L2 sentence comprehension has rarely been focused on (Hopp, 2015).

In the present study, based on accounts of visuo-linguistic integration (Clark and Chase, 1972) and situated language processing (Barsalou, 1999, 2008; Knoeferle, 2016; Knoeferle and Crocker, 2006), we assume that both visual and morphosyntactic information contribute to shaping mental representations. This study aims to test the robustness of visual effects on comprehension outcomes, and whether such effects vary between L1 and L2 speakers. We further explore the possibility that morphosyntactic sensitivity (i.e., the ability to interpret morphosyntactic cues to derive meaning) might predict to what extent the same individual makes use of visual information in spoken sentence comprehension.

4.3   Background

While earlier, 'syntax-first' accounts of sentence comprehension have typically as-sumed that (morpho)syntactic information is temporally prioritized over non-syntactic cues (Coltheart 1999; Fodor 1983; Frazier and Fodor 1978; Frazier and Rayner 1982; Frie-derici 2002), constraint-based approaches have proposed that all types of linguistic cues are integrated as they become available to shape mental representations (e.g., Trueswell et al. 1994). In both accounts, however, all linguistic information is assumed to be eventually considered for interpretation. The present study remains agnostic as to the temporal prec-edence of one cue type over others and assumes that non-linguistic (visual) cues are con-sidered alongside linguistic cues for sentence interpretation. In the following we review the evidence supporting the latter assumption. We then provide an overview of prior work on

the processing of noncanonical sentences by L1 and L2 speakers, focusing on ap-proaches which model conflicts between unambiguous syntax and other types of linguis-tic cues. Finally, we review evidence for the role of individual variability in cue-weighting patterns in L1 and L2 speakers.

4.3.1   The role of visual cues in L1 sentence comprehension

Language has increasingly been considered as naturally multimodal (especially in the spoken modality, Rosenblum, 2005, 2008), i.e., as immersed in and constantly interacting with the visual environment (for reviews, see Knoeferle, 2019; Knoeferle and Guerra, 2016). Early research on visuo-linguistic integration has typically used sentence–picture verification tasks to study speakers' preferences during language-mediated visual attention (Knoeferle et al., 2011, 2014; Underwood et al., 2004; Wannemacher, 1974). Early models of visuo-linguistic integration, such as the Constituent Comparison Model (CMM, Carpenter and Just, 1975) were put forth to account for response-time data from sentence–picture verification, and specifically for a facilitation in response to matching (i.e., 'true') relative to mismatching (i.e., 'false') trials (e.g., Clark and Chase, 1972). For instance, the CMM proposed that both the spoken and the visual components of the input are stored in terms of constituents and serially verified against one another. However, further research has painted a more complex picture.

For one, sentence–picture mismatches have been found to yield increases in verification accuracy. For example, in both experiments, Knoeferle et al. (2014) found that verification accuracy was higher for sentence–picture mismatches (vs. matches). This seems hard to reconcile with the CMM's claim that encountering a mismatching feature triggers the sentence–picture comparison process to be re-initiated (Carpenter and Just, 1975), as it is unclear how re-initiating the process may benefit verification success. Further, it has been debated whether a verification mechanism underlies situated language processing with no verbatim comparisons to be performed. For instance, Underwood et al. (2004) found no match–mismatch differences in sentence–picture verification times when visual scenes and written sentences were presented successively. This suggested that the process of extracting comparable components from visual and linguistic input for serial comparison may have been associated with the specific task settings in previous research (e.g., simultaneous presentation) and might not generalize. Knoeferle and Crocker (2005) addressed this issue by measuring word-by-word reading times of SVO/OVS sentences preceded by depictions that either matched or mismatched the sentences' thematic roles (e.g., SVO: Die Oma_{NOM/ACC} filmt soeben den_{ACC} Kaufmann 'The granny is filming the businessman right now'). Interestingly, however, the authors found selective slowdowns in reading times in mismatching trials, but no difference in total reading times (as in Underwood et al., 2004). This showed not only that a verification process was indeed taking place, though unrequired by the task, but also that mismatch effects were time locked to the specific cue that caused the sentence to mismatch

the depictions. The replication of Underwood et al.'s results further suggested that mismatch effects might only be measured with finer-grained temporal measures (vs. total response times), as such effects may be short lived. These results support the fundamental validity of claims from the CCM, contribute to extending its generalizability across tasks, and further point to the role of recency in modulating visual effects.

Visual cues have often been used as a means for measuring linguistic processing in the 'visual-world' eye-tracking paradigm (VWP), based on the assumption that oculomotor behavior and linguistic processing are coupled in a tightly time-locked fashion (e.g., Cooper, 1974). The VWP uses eye movements to visual displays to infer the time course and dynamics of the processing of particular cues during, e.g., word recognition (e.g., Allopenna et al., 1998; Mirman and Magnuson 2009) and incremental sentence processing (e.g., Altmann and Kamide, 1999, 2009; Kamide et al., 2003; Sedivy et al., 1999). For example, Kamide et al. (2003) examined the role of morphosyntactic and semantic cues in anticipating upcoming referents in unambiguous SVO/OVS sentences in German (e.g., OVS: Den Hasen frisst gleich der Fuchs 'The[ACC] hare[ACC] eats the[NOM] fox'). Participants were more likely to look at the fox when the first-mentioned NP 'the hare' was in the accusative (i.e., den Hasen, object role) in comparison to the nominative form (i.e., der Hase, subject role). This indicated that case-marking cues were integrated with lexical information and world knowledge to anticipate upcoming items.

Manipulating the visual input can also affect sentence processing (Tanenhaus et al., 1995), indicating a complex interaction of cross-modal information in sentence comprehension (Chambers et al., 2004; Knoeferle and Crocker, 2005, 2006). For example, Tanenhaus et al. (1995) showed that eye movements while processing temporarily ambiguous instructions (e.g., 'Put the apple on the towel in the box') were contingent on the visual display—for example, whether one or two apples were presented visually. Subsequent studies have provided ample evidence to support the claim that the integration of linguistic and visual cues plays a crucial role in determining language processing in context. For instance, Knoeferle et al. (2005) monitored eye-movements to visual scenes while participants listened to simple SVO/OVS sentences in German with initial thematic-role ambiguities (e.g., Die Prinzessin malt gleich den Fechter 'The princess[NOM/ACC] will soon paint the[ACC] fencer'). Crucially, the 'princess' character was depicted as though painting one character and being washed by another character. Participants were found to integrate morphosyntactic and visual cues incrementally to assign thematic roles at the verb, i.e., before the disambiguating case marker appeared on the second NP's determiner. Knoeferle et al.'s (2005) results provide strong evidence that visually encoded role relations are used to guide incremental interpretation of thematic roles in the face of syntactic ambiguity.

Yet, generalizing findings from studies with syntactically ambiguous stimuli to sentence processing at large may lead us to overestimate the role of non-syntactic cues for the processing of unambiguous structures (see, e.g., Folk and Morris, 2003; Hagoort, 2003; Hahne

and Jescheniak, 2001; Service et al., 2007). This is because ambiguous syntax allows for two or more interpretations, at least up to a point, such that participants may need to rely on non-syntactic cues to derive unambiguous mental representations. Thus, it may be hypothesized that non-syntactic cues are relied upon less if they are unnecessary for comprehension. Addressing this issue, Knoeferle and Crocker (2006, Experiment 2) presented unambiguous structures, which allows for a more conservative estimation of the contribution of non-syntactic cues. Using a similar design to Knoeferle et al. (2005), Knoeferle and Crocker (2006, Exp. 2) recorded participants' eye movements as they listened to unambiguous OVS structures in German and inspected depictions of three referents carrying out actions. The results showed that the depicted characters were attended to as soon as they were identified as potentially upcoming agents, extending the role for visual-scene information to the processing of syntactically unambiguous stimuli.

Building on and extending the CCM, the Coordinated Interplay Account (CIA, Knoeferle and Crocker, 2006) was proposed as a model of visuo-linguistic integration able to capture findings from recent research on situated language processing. The CIA defines the time course and boundary conditions of visuo-linguistic integration by describing a cyclic mechanism in two steps. First, as utterances are understood, comprehenders direct their visual attention towards the entities mentioned, to bind them to their referents and ground them in the real world. Second, the entities that are being visually attended to are integrated with spoken-language cues and contribute to the interpretation of the utterance.

The current study fundamentally assumes the premises of the CIA and seeks to ex-tend the evidence supporting it. Different from previous research, we measured the effects of conflicting (instead of compatible) visual-scene information on post-sentence comprehension, using unambiguous (instead of ambiguous) spoken sentences and a linguistic task unrelated to sentence–picture congruence. Sentence–picture verification tasks do not seem to be well suited to investigating the weighting of visual cues in shaping mental representations, because the visual input must be considered to verify a match or a mismatch. Thus, visual information can be assumed to contribute as much as linguistic information towards verification outcomes. Similarly, the VWP is often uninformative about how conflicting visual and linguistic cues contribute to shaping mental representations. This is because the conflict being modeled is often induced by syntactic ambiguity rather than by competing cues from distinct sources, as the visual displays are typically compatible with one or more possible interpretations. Similar to Knoeferle and Crocker (2006), we presented sentences with unambiguous thematic roles assigned to referents. We reasoned that, if visual influence on sentence processing is robust, effects of the visual context should emerge even when morphosyntactic cues suffice to derive fully specified meaning representations. Evidence for the integration of task-unrelated visual scenes on the comprehension of unambiguous

sentences would strengthen claims regarding the robustness of visual influence on linguistic processes.

4.3.2   Syntactic and non-syntactic cues in L1 vs. L2 sentence comprehension

The question as to the relative weighting of distinct linguistic cues has been the focus of the Competition Model (CM, Bates et al., 1982; Bates and MacWhinney, 1982; MacWhinney et al., 1984) and the dual-route approach (e.g., Ferreira, 2002). Similar to the present study, both of these accounts have typically inferred patterns of cue weightings by pitching cues of different types against one another, usually using agent-identification tasks.

The CM is a constraint-based model of language acquisition and processing. It presupposes that sentence interpretation is based on functional relations between form and meaning that are learnt probabilistically and can be implemented in a connectionist network. The model was put forth specifically to account for thematic-role-assignment preferences and introduced computable dimensions of linguistic cues. For instance, the dimension of cue reliability is computed as the proportion of times a cue leads to correct thematic-role assignment over its total occurrence. Crucially, cue dimensions are thought to take on language-specific values (or 'weightings'), based on which types of cues are prioritized for sentence interpretation. For example, MacWhinney et al. (1984) tested L1 English, German and Italian sentence comprehension while manipulating word-order, agreement, animacy and stress cues. The results showed that speakers' reliance on each type of cue varied systematically across languages. English L1 speakers used mainly word order, Italian L1 speakers relied mostly on agreement cues, and German L1 speakers attended to both agreement and animacy. The results were interpreted as evidence for cross-linguistic differences in cue weighting. Similarly, Kempe and MacWhinney (1999) manipulated case-marking and animacy cues to test thematic-role assignment using a speeded picture-choice task with L1 German and L1 Russian speakers. The reaction-time data showed that German but not Russian speakers relied on animacy cues, whilst Russian speakers relied on case marking to a larger extent than German speakers did. This suggests that case-marking cues are relied on more if they are more reliable for thematic-role assignment (as in Russian vs. German), although the inflectional paradigm is more complex in Russian (vs. German). Of relevance for our study is the authors' finding that less reliance on the morphosyntactic cue was associated with stronger reliance on the non-syntactic cue.

The CM has been extended to L2 sentence interpretation with the hypothesis that L1 cue weightings are transferred to L2 comprehension, especially at low levels of proficiency. For instance, Kempe and MacWhinney (1998) found that L1 English speakers learning Russian relied on Russian case markings earlier than L1 English speakers learning German relied on German case markings. Instead, the learners of German tended to consider both animacy and

case cues, even though the latter unambiguously indicated the sentence participants' thematic roles. These findings strongly suggest that differences in cue weighting might underlie L1 vs. L2 comprehension strategies. However, the experimental setup in CM studies often resulted in the presentation of ungrammatical stimuli alongside (semi-)grammatical ones (Bates and MacWhinney, 1989). This design choice has been sharply criticized as it may influence participants' response strategies, or be confusing to participants, thus limiting the conclusions to be drawn from these designs (e.g., Gibson, 1992). Further, due to the large number of conditions resulting from the manipulation of several cues at once, and to relatively small sample sizes, CM studies have been systematically underpowered.

Effects of conflict between syntactic and non-syntactic cues on comprehension accuracy have also been the focus of dual-route approaches to sentence comprehension. Within this research, comprehenders have been found to mis-assign the thematic roles of non-canonical sentences (Karimi and Ferreira, 2016). Such misinterpretations are facilitated when non-syntactic cues, such as semantic plausibility, support the incorrect reading. Linear word-order heuristics underlie the 'agent-before-patient' preference, by which speakers overwhelmingly interpret Noun–Verb–Noun (NVN) structures as Subject–Verb–Object (SVO, e.g., Schriefers et al., 1995; Ferreira, 2003). For instance, Ferreira (2003, Experiment 3) examined the comprehension of unambiguous English sentences containing subject and object clefts as a manipulation of linear word order (e.g., subject cleft: It was the dog that bit the man). Pragmatic plausibility was manipulated to convey plausible and implausible readings (e.g., implausible object cleft: It was the dog the man bit). The results showed that agent-identification accuracy was higher for subject clefts overall, and the effect of implausible (vs. plausible) relations was evident in participants responses to object (but not subject) clefts. To account for this and further evidence, comprehenders have been argued to construe meaning either based on the algorithmic parsing of syntactic cues, including case markings, or on heuristics relying on non-syntactic information such as plausibility and word-order cues (Christianson, 2016; Ferreira, 2002). Thus, especially in the face of noncanonical syntactic structures, comprehenders may endorse interpretations that are inconsistent with the linguistic input as computed algorithmically (Ferreira and Patson, 2007; Karimi and Ferreira, 2016). The core ideas of the dual-route approach can be re-mapped onto a multiple-constraint framework in which distinct cues are assumed to take on different weightings. The present study adopts an analogous rationale to the dual-route approach in that it models situations in which morphosyntactic and visual cues convey competing interpretations, but without assuming a dichotomous division of (algorithmic vs. shallow) routes to interpretation (e.g., Kuperberg, 2007).

Relative to English, German has a more flexible word order due to the use of morpho-logical case markings. Thus, German linear word order is not a reliable cue, because an NVN structure cannot always be mapped onto an SVO structure (as in English). Case markers may

provide more reliable indicators of thematic roles, allowing for the argument phrases that carry them to be mapped onto a verb's thematic grid. In particular, masculine case markings present unambiguous nominative vs. accusative cues (der vs. den) for grammatical role assignment, which are used for prediction (Kamide et al., 2003) and may eliminate the 'agent-before-patient' preference observed for German ambiguous NVN clauses (e.g., Schriefers et al., 1995). Yet, L1 German speakers have also been shown to have problems interpreting unambiguous, noncanonical sentences in agent-identification tasks (Bader and Meng, 2018; Meng and Bader, 2021). For example, Meng and Bader (2021) found that agent-identification accuracy rates significantly decreased in response to OS (vs. SO) sentences, regardless of whether the thematic roles were reversible (e.g., *The father hugs the uncle*), biased (e.g., *The chef ruined the roast*), or non-reversible (e.g., *The chef cleaned the pan*). These results suggest a crucial role of linear word order (giving rise to the agent-before-patient preference) for agent-identification success. Therefore, in the pre-sent study, word order is taken to represent a further heuristic for German sentence interpretation, one that will only be beneficial for the interpretation of more frequent, canonical structures. We manipulated the conflict between linear word-order-based thematic role biases and case cues by manipulating the relative ordering of nominative- and accusative-marked nominals.

As regards L2 comprehension, the Shallow Structure Hypothesis (SSH, Clahsen and Felser, 2006, 2018) applies a similar rationale to the dual-route approach to understanding L1–L2 differences in sentence processing. The SSH was proposed to account for evidence showing that L2 morphosyntactic processing tends to be less automatized compared to L1 processing (Clahsen and Felser, 2006, 2018). It is supported by findings showing persisting vulnerability in the use of morphosyntactic cues in an L2, which may be compensated for by an increased reliance on non-syntactic relative to syntactic information (e.g., Cunnings et al., 2017; Deniz, 2022; Felser and Cunnings, 2012; Felser et al., 2003; Foucart et al., 2015; Lee and Witzel, 2022; Mitsugi and MacWhinney, 2016; Pan and Felser, 2011; Pan et al., 2015; Papadopoulou and Clahsen, 2003; Pozzan and Trueswell, 2016). Cunnings's (2017) Interference Hypothesis is similar in spirit but holds that L1–L2 differences in cue weightings specifically affect memory retrieval during processing. Support for differential cue weighting in L1–L2 comprehension is provided by studies exploring the extent to which L2 speakers at different proficiency levels integrate information from morphosyntax and, e.g., prosody (den Ouden et al., 2016), pragmatic consistency (Foucart et al., 2015), semantic plausibility (Jacob and Felser, 2016; Roberts and Felser, 2011), discourse context (Leal et al., 2017; Pan and Felser, 2011; Pan et al., 2015; Puebla and Felser, 2022) or focus (Patterson et al., 2017). However, little is known about the impact of visual information on L2 sentence comprehension.

4.3.3 *Individual variability in L1 and L2 sentence comprehension*

While most previous research on sentence comprehension has analyzed group-level performance, essentially treating individual variability as a source of noise, there has been growing recognition of the fact that individual variability may be an important source of information to consider as well. Individual differences in L1 sentence comprehension have often been considered in terms of working-memory capacity (e.g., King and Just, 1991; Pearlmutter and MacDonald, 1995) and reading speed (e.g., Yadav et al., 2022). Results from this line of research demonstrate that individual differences in cognitive capacity or linguistic abilities might determine sensitivity to different information sources in L1 comprehension. However, our understanding of how individual-difference variables affect L2 performance is still rather limited (see Farmer et al., 2012; Kidd et al., 2018, for review and discussion). While previous L2 research has often investigated the role of L2 proficiency, language background, age of acquisition, or cognitive or affective variables, few studies have examined variables related to individual differences in specific linguistic abilities (e.g., Hopp, 2015; Kilborn, 1992; Roberts and Felser, 2011). For instance, Hopp (2015) measured speakers' ability to integrate semantic and syntactic information using a word-monitoring task. Interestingly, semantic- and syntactic-integration abilities were found to correlate negatively in L2 speakers, suggesting that the larger the advantage experienced by each participant in the presence (vs. absence) of semantic cues, the smaller the advantage experienced in the presence (vs. absence) of syntactic cues, and vice versa. This result seems compatible with the dual-route and SSH claims regarding a potential trade-off in the reliance on syntactic vs. non-syntactic cues.

Indirect support for a dual-route approach to sentence comprehension, as well as novel insights into individual variability, were brought about by electrophysiological findings showing systematic differences in speakers' brain responses to grammatical and semantic violations at the individual level. These were interpreted as reflecting individual tendencies towards reliance on algorithmic versus superficial processing in L1 speakers (Osterhout, 1997; Pakulak and Neville, 2010). Further studies have shown that the brain signatures of L2 speakers may show an increased reliance on a shallow processing strategy, and that increasing grammatical sensitivity is associated with increasing overlap with L1 speakers' brain signatures (Morgan-Short et al., 2010; Tanner et al., 2014; Tanner et al., 2013). This evidence suggests that graded differences in sensitivity to different cues may underlie individual patterns of linguistic performance beyond categorical grouping.

Summarizing, individual variability has been observed in both L1 and L2 comprehension, both at the behavioral and the neurophysiological level, but the degree to which different kinds of individual-difference measures can predict L1 and L2 performance patterns is still far from clear.

4.4   Aims and research questions

This study focused on the interplay of visual-scene and morphosyntactic cues for thematic-role assignment of canonical and noncanonical sentences in German L1 and L2 comprehension. Specifically, we investigated the following research questions:

1. How do (task-unrelated) visual cues influence thematic-role assignment in unambiguous sentences in L1 and L2 speakers?

2. Does individual sensitivity to morphosyntactic cues predict speakers' sensitivity to visual cues?

The aim of Research Question 1 is to provide evidence for the robustness of visual-scene effects (as predicted by the CIA) in a maximally conservative setting, and to extend previous evidence on situated language to L2 speakers. It also explores the claim that that (morpho)syntactic and non-syntactic cues are differently weighted in L1 and L2 comprehension (e.g., Clahsen and Felser, 2006; 2018; Cunnings, 2017), extending previous research by investigating the relative impact of visual context information on comprehension. Re-search Question 2 builds on the SSH/dual-route claim for a potential trade-off in the reliance on syntactic and non-syntactic cues, and focuses on the role of individual differences in L1 and L2 sentence comprehension. Considering that traditional group-level analyses may mask graded differences in cue weightings between individuals (see previous Section), we test (i) whether reliance on syntactic and visual information are negatively as-sociated at the participant level, and (ii) whether individual cue weighting can explain comprehension performance beyond categorical L1–L2 grouping.

To answer Research Question 2, we need to obtain a proxy of each participant's sensitivity to morphosyntactic cues. Experiment 1 sought to replicate previous findings on the L1 and L2 comprehension of noncanonical sentences and to provide an individual index of sensitivity to use as a predictor in Experiment 2.

Here, we focus on morphological case marking as the syntactic cue under investigation. While case has been found to be less problematic in L2 acquisition and processing compared to, for example, agreement or tense marking (e.g., Lardiere, 1998; Sato and Felser, 2010), L2 comprehenders have shown reduced sensitivity to case cues compared to L1 comprehenders (e.g., Jackson, 2008; Jackson and Bobb, 2009; Mitsugi and MacWhinney, 2016 - but cf. Schlenter and Felser, 2021).

We chose L2 speakers whose L1s present a strong case-marking system so as to minimize the possibility of negative cross-linguistic influence on the decoding of case information. Because the L1s of the participants included in the L2 group rely on case marking for grammatical role assignment, similarly to German (Arnaudova et al., 2004; Benţea, 2012; Citko, 2016; Levy et al., 2013; Sevcenco et al., 2013), any patterns signaling that L2 speakers

weight morphosyntactic cues less heavily than L1 speakers cannot be easily attributed to the absence of case-marking cues in the L1.

4.5   Methods

4.5.1   Design

In two experiments, we manipulated the structure of unambiguous spoken relative clauses in German (subject-extracted and object-extracted relative clauses; SR and OR henceforth), which were presented in isolation (Experiment 1) and alongside scene depictions (Experiment 2). Scene depictions were manipulated either to convey the same (match) or the opposite (mismatch) role relations as the spoken sentence. Participants answered a forced-choice binary agent-identification question referring to the spoken sentence. Accu-racy and response latencies were analyzed using mixed-effects logistic and linear regression models. All participants took part in both experiments.

4.5.2   Participants

Participants were recruited via the online participant recruitment platform of the University of Potsdam and via web-based adverts throughout Germany. Participants provided their informed consent to participating in this study and were compensated with either 10 EUR/h or university credit. Participants reported no perceptual, cognitive, or language-related impairments, and normal or corrected-to-normal vision (including un-impaired color vision).

Our L1 participants were speakers of German who had not learned another language before the age of 5. In our L2 group, we recruited L1 speakers of Balto-Slavic languages with a strong case-marking system (i.e., Bulgarian, Czech, Latvian, Lithuanian, Polish, Russian, Slovakian, Ukrainian) as well as Romanian speakers. What we refer to as strong case-marking system is one that does not only operate on pronouns, a, e.g., in English. Romanian was included as an eligible L1 because, although it does not belong to the Balto-Slavic languages, it differs from other Romance languages in that it has a case system comparable to that of German (Benţea, 2012).

Our sample included 84 participants from 19 to 50 years of age (mean = 28, SD = 8), of which 45 L1 and 39 L2 speakers of German. The same participants took part in both Experiments. Of the L2 speakers, one was scored as A2, 13 as B1, 18 as B2 and seven as C1 on the Common European Framework of Reference for Languages (CEFR) scale. For all analyses, A2 and B1 speakers were collapsed into the same proficiency group. Further participant characteristics are reported in Table 1.

Table 1: Participant information. Demographic information is reported for both groups under 'Demographic information'. Under 'L2 group', we report the L2 speakers' background of German learning and percentage of daily use. Under 'Additional language information', we report information on L1 and additional languages spoken in both groups.

Demographic Information					
Group	Female	Male	Total	Mean Age in Years (SD)	Mean Years of Education (SD)
L1	36	9	45	24 (4)	15 (2)
L2	35	4	39	32 (9)	16 (5)
Total	71	13	84	28 (8)	16 (4)

L2 Group							
Age of Acquisition of German	N	Proficiency in German	N	German Learning Context	N	Daily use of German	N
3 to 5 yrs	6	C1	7	School	35	100%	7
6 to 10 yrs	10	B2	18	Home	3	90%	5
11 to 15 yrs	10	B1	13	Informal	1	80%	10
16 to 20 yrs	8	A2	1			70%	4
21 to 25 yrs	5					60%	5
26 to 30 yrs	1					50%	3
31 to 35 yrs	0					40%	0
36 to 38 yrs	1					30%	1
						20%	3
						10%	1

Additional Language Information					
L1	N	Spoken Languages	L1 Group	L2 Group	Additional Languages (N)	Mean AoA in Years (SD)
Bulgarian	4	L1	2	0	English (74)	7 (4)
Czech	4	L1 + 1	14	8	French (20)	7 (6)
German	45	L1 + 2	21	18	Spanish (9)	10 (4)
Latvian	1	L1 + 3	7	4	Russian (7)	6 (5)
Polish	5	L1 + 4	1	7	Italian (3)	9 (4)
Romanian	8	L1 + 5	0	2		
Russian	15					
Slovakian	1					
Ukrainian	1					

4.5.3   Materials

4.5.3.1   Spoken sentences (Experiments 1 and 2)

The linguistic stimuli included 80 spoken sentences recorded by a female speaker of German. They were identical for Experiments 1 and 2. Target sentences were distributed across two (Experiment 1) or four (Experiment 2) presentation lists using a Latin-Square design to counterbalance the experimental conditions across participants. Each presentation list contained eight practice items, 24 targets, 16 pseudo-fillers and 32 fillers. Item order within each list was randomized for each participant. The complete presentation lists are available in the Supplementary Materials (https://osf.io/m5ujv/).

Target sentences were unambiguous copular sentences containing a modifying relative clause. The main clause introduced a human male referent denoted by a nominative case-marked NP. This was followed by a relative clause introducing a (human) female referent and

a transitive verb. See Table 1 for examples. All characters were common nouns indicating grammatically gendered referents (e.g., *die Braut* 'the bride'; *der Prinz* 'the prince') or professions (e.g., *der Koch* 'the cook'). For the manipulation of syntactic structure, half of the relative pronouns were in nominative (*der* 'who') and half in the accusative case (*den* 'whom'), introducing subject-extracted (SR) and object-extracted (OR) relative clauses, respectively. Only masculine nouns were included as antecedents because masculine singular pronouns allow for unambiguous case marking in German. The argument order in SRs represents the canonical word order in German (NOM > ACC), whilst ORs are noncanonically ordered (ACC > NOM). To interpret OR sentences correctly, listeners need to associate the main clause subject (e.g., *der Koch* 'the cook') with the embedded verb's (e.g., *verfolgt* 'follows') object or patient role. An accusative-marked relative pronoun provides an unambiguous cue to this interpretation.

Pseudo-fillers were identical to target items, except that they included feminine, neuter, or plural relative pronoun antecedents. All items were globally unambiguous, but some fillers presented local syntactic ambiguity due to case syncretism. Such ambiguities were always resolved by subject-verb agreement or unambiguous case marking inside the relative clause, such that all stimuli were globally unambiguous, and participants did not need visual cues to solve the task (e.g., pseudo-filler sentence: *Die sind die Damen, die der Dieb erschießt* 'These are the ladies whom the thief shoots dead'; probe question: *Wird der Dieb erschossen?* 'Is the thief shot dead?'). Locally ambiguous fillers were introduced to prevent habituation, given that all target items contained masculine relative pronouns. To further increase variety, pseudo-fillers also included non-human animate referents (e.g., *Hund* 'dog,' *Pferd* 'horse') and inanimate referents (e.g., *Vase* 'vase,' *Tisch* 'table'). Pseudo-fillers included an equal number of subject-relative and object-relative clauses to replicate the symmetrical distribution of sentence structures in the critical items. Filler sentences included a main clause and a causal or temporal subordinate clause. In half of the fillers, the subordinate clause preceded the main clause, while in the other half, it followed it (e.g., *Der Hund verfolgt den Mann, weil er mit ihm gehen möchte* 'The dog follows the man be-cause it wants to walk with him'; *Weil sie einen Schatz gefunden haben, freuen sich die Kinder sehr* 'Because they found a treasure, the children are very happy').

4.5.3.2   *Picture stimuli (Experiment 2)*

Scene depictions consisted of photos of Playmobil© figurines performing actions against a white background. The photos were taken in a brightly lit room with a Canon EOS 40D digital camera and edited with the open-source software GIMP (GNU Image Manipulation Program) software (version 2.10.20; The GIMP Development Team, 2019). All pictures were cropped to a 1:1 width-to-height ratio, and their visual features were kept as similar as possible to one another (e.g., characters' size and location on the screen, their respective position, brightness, and saturation of color features).

Each image depicted only the two referents mentioned in the sentence. For target sentences, visual cues in mismatching conditions deviated from the meaning of the sentence only with regard to the conveyed role relations, i.e., the action direction, with half of the pictures matching the thematic roles in the auditory stimulus sentence (match condition), and the other half presenting them as reversed (mismatch condition). See Table 2 for examples of target pictorial stimuli.

Table 2: Experimental design and sample stimuli. NOM = nominative case, ACC = accusative case.

First Slide				Second Slide
Sentence Structure		Visual Cue		Written Probe Question
Level	Example	Level	Example	
Subject relative (SR)	*Das ist der Koch, der die Braut verfolgt.* this is theNOM cook, whoNOM theACC bride follows 'This is the cook who is following the bride.'	Match		*Wird die Braut/der Koch verfolgt?* is theNOM bride/theNOM cook followed 'Is the bride/the cook being followed?'
		Mismatch		
Object relative (OR)	*Das ist der Koch, den die Braut verfolgt.* this is theNOM cook, whoACC theNOM bride follows 'This is the cook whom the bride is following.'	Match		
		Mismatch		

As was the case for target items, fillers and pseudo-fillers were presented half with matching and half with mismatching depictions. In order to dissuade participants from paying attention specifically to the thematic roles depicted in the image, context mismatches for fillers and pseudo-fillers were (in equal parts) due to either the verb (i.e., picture and sentence described two different actions), locative prepositional phrases (i.e., the sentence stated that an object was in a different position from how it was depicted), or a number mismatch (i.e., the sentence stated that a different number of referents were present than were depicted).

4.5.3.3   Probe questions (Experiments 1 and 2)

Each sentence was directly followed by a slide presenting a written probe question (in the passive voice) which targeted agent identification. Half of the probe questions required a *Yes*

and half a *No* response. See Table 2 for example questions. Half of the questions referred to the male and half to the female character. Thus, responding accurately required participants to deeply parse the target clause structures, as the most reliable cue was case marking information on the relative pronoun. Although errors in delayed agent-identification tasks may be associated with memory failures rather than parsing difficulty (Meng and Bader, 2021), presenting probe questions targeting the processing of case-marking information has been shown to encourage speakers' reliance on morphosyntactic cues (Swets et al., 2008).

The experiment was programmed in E-Prime 3.0 (Psychology Software Tools 2016, Pittsburgh, PA) and conducted online using E-Prime Go (Psychology Software Tools 2020, Pittsburgh, PA). All participants completed an online screening questionnaire and then received an e-mail with instructions, a download link for the experiment, and a subject ID number. The order of tasks was arranged in two sentence-comprehension blocks (Experiments 1 and 2) in counterbalanced order.

4.5.4 Procedure

4.5.4.1 Screening questionnaire

Participants filled out a demographic background questionnaire, asking about their gender, date of birth, number of years of education, L1, other spoken languages (for each, specifying age-of-acquisition and acquisition context), handedness, vision and/or hearing deficits, and (past or current) language and/or neurological impairments. If participants were L2 speakers of German, they were additionally asked to indicate their age of acquisition of German, the context in which they learned German, and the extent to which they used German in their daily life on a percentage scale. The language history section of the questionnaire for L2 participants followed the guidelines outlined in Marian et al. (2007).

4.5.4.2 German proficiency assessment

All L2 participants were asked to complete an online test to measure their proficiency level in German (https://www.sprachtest.de/einstufungstest-deutsch, accessed on 19 November 2021). The test is a freely available online tool containing language tasks aimed at assessing reading- and listening-comprehension skills, as well as grammar and vocabulary. The test is timed, and its duration is 15 minutes. The evaluation of the participants' skills is based on the points obtained for each category: reading (max. 5 points), listening (max. 7 points) and grammar and vocabulary (max. 28 points; total = 40 points). Participants' proficiency levels were derived automatically from the raw scores according to the Common European Framework of Reference's guidelines (Council of Europe, 2011).

4.5.4.3   Comprehension task

Each trial in the sentence-comprehension tasks consisted of two steps: first, participants heard a spoken sentence; then, a written binary forced-choice probe question was presented which participants were asked to answer through keypress within 4000ms. As mentioned above, the task was administered in two consecutive blocks. Experiment 1 presented sentences in isolation and Experiment 2 presented sentences alongside matching/mismatching visual cues. The order in which both experiments were administered was counterbalanced across participants. Participants were allowed to take a break between both experiments. Target items were presented in a Latin-Square design, such that each participant only saw target items once within and across both experiments. In contrast, filler and pseudo-filler sentences were presented identically twice, though the correct response was counterbalanced across experiments to prevent habituation. This was achieved by altering the probe question. Participants completed eight practice trials before each experiment. Of the practice trials, four resembled target and pseudo-filler items, and four resembled filler items. Practice trials were distributed symmetrically across conditions: half of target-like trials presenting subject-relatives and half object-relative clauses; in addition, in Experiment 2, both filler- and target-type practice trials were presented alongside half matching, half mismatching depictions. Practice probe questions required an equal number of Yes and No responses. All items were presented to participants in randomized order. The total duration of the procedure was about 25 min.

4.5.5   Data analyses

All analyses were conducted in R Studio (version: 2021.09.0, RStudio Team, 2021). The analysis code is available in the Supplementary Materials. All trials to which no response had been provided and with response times below and above two standard deviations from the mean were removed from the dataset. This operation resulted in the exclusion of all trials from two participants and altogether 419 trials, corresponding to 9.59% of the whole dataset. Further, one participant had to be excluded from Experiment 2 because they provided no responses to Experiment 1.

We collected accuracy rates and RTs (for correct responses) in response to the probe questions. We fitted mixed-effects logistic regression models (binomial family) for analysis of accuracy and linear mixed-effects regression models for RTs (natural log). Mixed-effects logistic and linear regression analyses were carried out with the lme4 R package (Bates et al., 2015). The effects of individual-level variables (e.g., proficiency, age of acquisition) on sensitivity scores were assessed with simple linear regressions. Models were plotted using *sjPlot* (Lüdecke, 2018) and *ggplot2* (Wickham, 2016). Follow-up analyses were performed with the _emmeans_ package (Lenth, 2022).

Sentence structure (SR, OR), visual cue (match, mismatch), and group (L1, L2) were sum-coded with ±0.5 weights, which allows us to compare the effect of an OR (0.5) relative to an SR clause (−0.5), the effect of a mismatch (0.5) relative to a match (−0.5), and the effect of belonging to the L2 group (0.5) compared to the L1 (−0.5). Note that the intercept represents the grand mean, i.e., the average response outcome across conditions for the whole sample.

Accuracy rates in Experiment 1 were averaged at the individual level and used as proxies of sensitivity to morphosyntax ('sensitivity scores' henceforth), as high sensitivity to case information would result in higher accuracy overall. This is because participants who are not sensitive to case marking but rather rely on surface-level heuristics (such as word order and pattern frequency), will assign thematic roles correctly only 50% of the time. Kempe and MacWhinney (1998) followed a similar rationale to estimate "case-marking mastery" (p. 560) and used it as a predictor for further analyses.

To answer Research Question 1, we fitted a mixed-effects logistic regression (binomial family) to accuracy rates from Experiment 2, which included sentence structure, visual cue, and group as fixed factors. To answer Research Question 2, we fitted an identical model but replaced the group factor with continuous sensitivity scores (centered). In both latter models, the random-effects structure included by-item and by-participant intercepts, and by-participant slopes for sentence structure and visual cue. Finally, we compared both models with ANOVA comparisons. This allowed us to assess whether sensitivity scores represent a better predictor for thematic-role assignment success than L1–L2 grouping.

4.6   Experiment 1

4.6.1   Predictions

4.6.1.1   Replication of subject preference and L1 advantage

On the group level, we expect canonical SR (vs. OR) clauses to yield higher accuracy and shorter reaction times (RTs) across both groups. This is based on findings for less ac-curate agent-identification outcomes for German object- (vs. subject-) extracted relative clauses (Bader and Meng, 2018, 2023; Meng and Bader, 2021). We assume that the subject preference reflects an overwhelming reliance on an agent-before-patient strategy (see Ferreira, 2003; Karimi and Ferreira, 2016), facilitated through the considerably higher occurrence frequency of SO vs. OS order in German (e.g., Bader and Häussler, 2010). We use the subject preference as a testbed to evaluate predictions from dual-route approaches for both L1 and L2 speakers. Specifically, we explore whether visual cues modulate the ease with which unambiguous syntax can successfully be interpreted in sentences with canonical and noncanonical ordering of thematic roles.

Overall, L1 speakers are expected to respond faster and more accurately than L2 speakers overall, based on evidence for L1–L2 differences in sensitivity to case information (e.g., Jackson, 2008; Jackson and Bobb, 2009; Mitsugi and MacWhinney, 2016; Rankin, 2014). Specifically, if L2 speakers are less able to make use of case information (i.e., have low morphosyntactic sensitivity), they might show disproportionate accuracy decreases in response to OR (vs. SR) clauses, in line with the SSH. This is because correct thematic-role assignment in SR trials may either be performed via deep syntactic parsing or via surface-level heuristics, as both routes convey the same reading. In contrast, in response to OR clauses, successful thematic-role assignment can only happen through the parsing and integration of morphosyntactic cues into a mental representation.

4.6.1.2   *Individual proxies of sensitivity to morphosyntax*

We expect L1–L2 differences in morphosyntactic sensitivity scores, which should be significantly higher in L1 than L2 speakers overall, as indicated by the effect of group on accuracy scores in Experiment 1. Increasing L2 proficiency might be associated with higher sensitivity to morphosyntactic cues. Further, based on previous evidence in favor of a role for maturational processes in the mechanisms underlying both morphological (Veríssimo et al., 2018) and syntactic processing (Pakulak and Neville, 2011; Díaz et al., 2016), increasing age of acquisition of German might be associated with lower sensitivity to morphosyntactic cues.

4.6.2   Results

4.6.2.1   *Descriptives*

Average proportions of accurate responses and RTs in milliseconds to correctly answered trials from Experiment 1 are reported for each group and condition in Table 3. While accuracy rates were numerically similar in L1 and L2 speakers in response to SR clauses, we noticed larger group differences in response to OR clauses, as well as higher variability.

Table 3: Average proportions of accuracy and RTs in milliseconds to correctly answered trials for L1 and L2 speakers.

Group	Sentence Structure	Accuracy	SD Accuracy	Reaction Times	SD Reaction Times
L1	SR	0.88	0.32	1,112	410
L1	OR	0.68	0.47	1,219	471
L2	SR	0.85	0.36	1,318	491
L2	OR	0.45	0.50	1,369	541

4.6.2.2   *Replication of group-level subject preference and L1 advantage*

The models fitted to accuracy and response times included random by-item and by-participant intercepts. The logistic regression fitted to accuracy data also included by-participant sentence-structure slopes, which were excluded from the RT model due to convergence errors. The coefficients of the models are reported in Table 4.

Table 4: Mixed-effects regression coefficients from linear and logistic regressions fitted to the natural log of response times of correctly answered trials and accuracy rates. sentence structure, visual cue, and group were sum-coded with ±0.5 weights, comparing the effect of OR (0.5) versus SR clauses (−0.5), of mismatch (0.5) versus match (−0.5), and of pertaining to the L2 (0.5) versus the L1 (−0.5) group. The intercept represents the grand mean.

Reaction Times			
Predictor	Estimate	Std. Error	t Value	p Value
(Intercept)	7.07	0.03	219.59	<0.001
Sentence Structure	0.07	0.02	3.71	<0.001
Group	0.15	0.06	2.55	0.011
Sentence Structure × Group	−0.03	0.04	−0.75	0.455
Accuracy			
Predictor	Estimate	Std. Error	t Value	p Value
(Intercept)	1.26	0.12	10.40	<0.001
Sentence Structure	−1.83	0.22	−8.36	<0.001
Group	−0.82	0.21	−3.85	<0.001
Sentence Structure × Group	−0.86	0.41	−2.08	0.038

Formulae: log(RT) ~ sentence structure × group + (1 | subject) + (1 | item). Accuracy ~ sentence structure × group + (sentence structure | subject) + (1 | item).

The model fitted to the RT data showed the main effects of sentence structure, with SR clauses yielding faster responses than OR clauses, and of group, as L2 speakers responded more slowly than L1 speakers. We found no evidence for an interaction of both factors. The model fitted to the accuracy data revealed a main effect of sentence structure, showing that accuracy was higher in response to SR than OR clauses. The main effect of group shows that L2 speakers' accuracy was lower overall than the German controls'. Further, the significant interaction of sentence structure × group indicates that L2 comprehension accuracy was reduced specifically for OR clauses, as compared to L1 speakers (Figure 1). Post hoc tests revealed significant between-group differences in accuracy responding to OR ($b = 1.25$, $SE = 0.29$, $z = 4.27$, $p < 0.001$) but not SR clauses ($b = 0.39$, $SE = 0.30$, $z = 1.30$, $p = 0.192$). These patterns confirmed our predictions by replicating previously reported accuracy decreases for OR clauses in German (e.g., Bader and Meng, 2018, 2023) and L1–L2 differences in the comprehension of noncanonical syntax (e.g., Mitsugi and MacWhinney, 2016; Jackson, 2008; Jackson and Bobb, 2009).
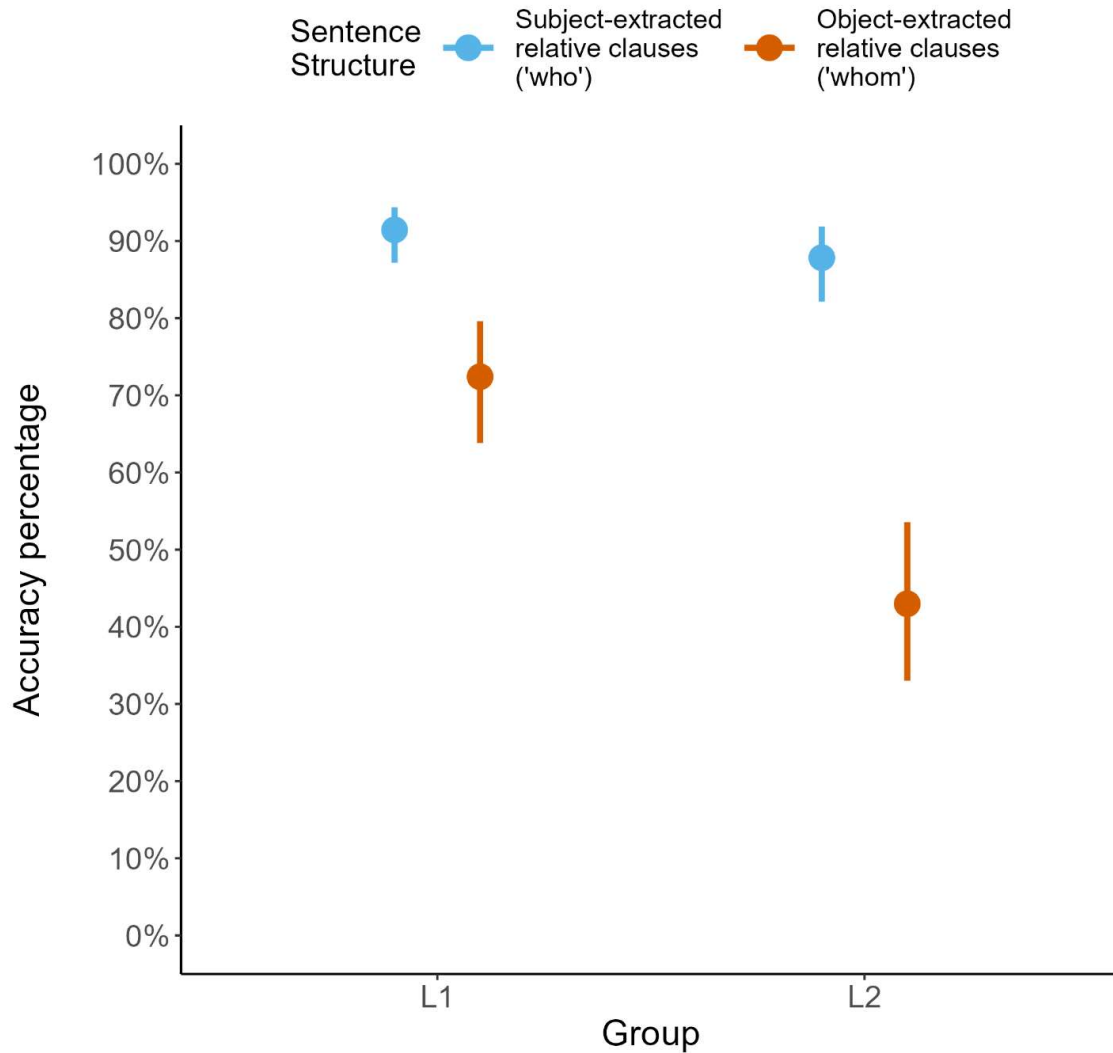
Fig 1 Interaction of sentence structure by group on accuracy rates in Experiment 1. Error bars represent 95% confidence intervals

4.6.2.3   Individual differences in sensitivity to morphosyntactic cues

By-participant sensitivity scores are plotted in Figure 2 as a function of group. After computing sensitivity scores, we examined their relation to individual-level variables with simple regressions. As suggested by the group analysis, belonging to the L2 (vs. L1) group was associated with lower sensitivity scores (b = −0.14, SE = 0.03, t = −4.28, p < 0.001). Proficiency in German contributed to determining sensitivity, which was significantly decreased in C1 (vs. L1) speakers (b = −0.46, SE = 0.11, t = −4.03, p < 0.001), and in B1 (vs. the average of L1, C1 and B2) speakers (b = −0.09, SE = 0.04, t = −2.09, p = 0.039). Moreover, increasing age of acquisition of German (centered) was strongly associated with lower sensitivity scores (b = −0.01, SE < 0.01, t = −4.18, p < 0.001). Both latter patterns are in line with previous evidence for the negative effect of increasing age of acquisition on the automatized ability to make use of subtle morphological and morphosyntactic cues to derive meaning (e.g.,

Veríssimo et al., 2018), which in turn is closely reflected by measures such as proficiency (e.g., Rossi et al., 2006).
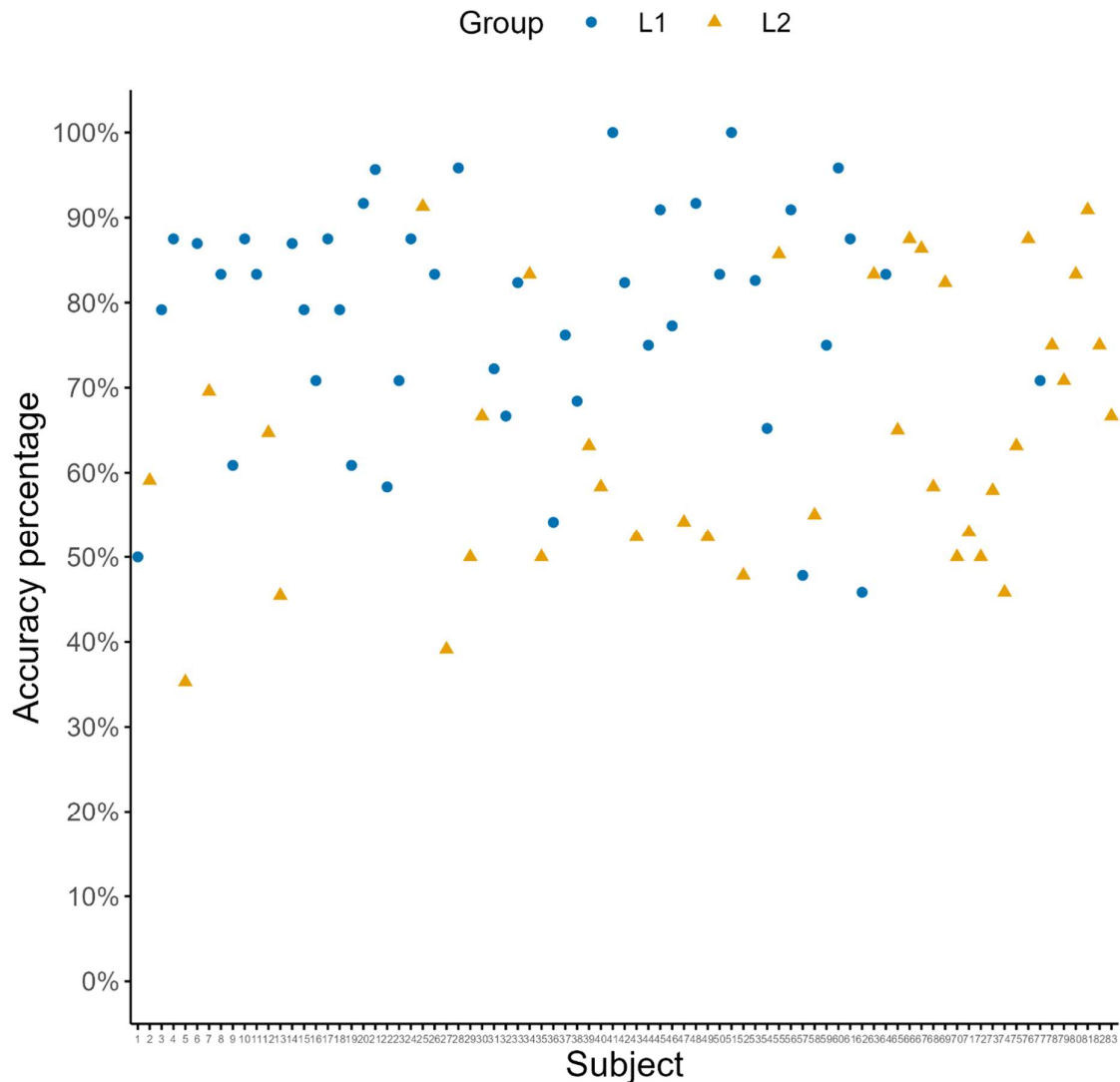
Fig 2 By-participant accuracy rates from Experiment 1 (i.e., sensitivity scores) as a function of group

To summarize, the results of Experiment 1 show that sentence structure influenced both dependent variables, with SR clauses being answered more accurately and faster than OR clauses across groups. This finding is compatible with evidence for decreased agent-identification accuracy of German OS (vs. SO) clauses (Bader and Meng, 2018, 2023; Meng and Bader, 2021). Further, the results confirmed that L2 speakers were slower and less accurate at comprehending the spoken sentences than L1 speakers, and their accuracy rates were especially low in response to OR clauses. The latter results replicate the L1 advantage and suggest that L2 speakers may have greater difficulty at comprehending non-canonical structures than L1 speakers. These conclusions are in line with previous evidence for a

decreased tendency to rely on morphosyntactic information to process the L2 (vs. L1, e.g., Grüter et al., 2020; Mitsugi and MacWhinney, 2016).

4.7   Experiment 2

4.7.1   Predictions

Overall, we expect to replicate the results from Experiment 1: SR (vs. OR) clauses should be understood faster and more accurately, and L1 speakers are expected to respond faster and more accurately than L2 speakers. L2 speakers may also show selective decreases in accuracy rates in response to OR (vs. SR) clauses. However, the sentence structure × group interaction detected in Experiment 1 may be modulated by the presence of visual cues (see below). As sensitivity scores were computed as the mean individual accuracy in Experiment 1, we expect such scores to be associated with more accurate performance in Experiment 2.

4.7.1.1   Research question 1: Visual influences on L1 and L2 sentence comprehension

We expect sentence–picture congruence (vs. incongruence) to yield faster and more accurate responses overall (i.e., across clause types and groups). This hypothesis is in line with evidence in favor of the interplay of visual and morphosyntactic cues in (i) the processing of ambiguous syntax with tasks unrelated to sentence–picture congruence (e.g., Knoeferle and Crocker, 2005), and (ii) the processing of unambiguous syntax (e.g., Knoeferle and Crocker, 2006; Knoeferle, 2007). We specifically add to Knoeferle and Crocker (2006) by examining whether visually driven representations are retained and influence comprehension post-sentence. Because the CIA predicts a role of recency in determining the strength of visual effects, these may be too weak to be detected on post-sentence comprehension (Knoeferle and Crocker, 2007). To increase the recency of the visual inspection relative to the probe question, we opted for simultaneous presentation of spoken and visual stimuli. In sum, the present setting enables us to test the robustness of visual influences across stimuli, tasks, and populations, and thus to extend the generalizability of models such as the CCM and the CIA.

In both speaker groups, the comprehension of noncanonical (vs. canonical) clauses may be influenced by visual cues to a larger extent. Specifically, for the comprehension of OR clauses, matching visual cues may make it easier for comprehenders to overcome the general subject preference, while mismatching visual cues may render it more difficult. This is because in mismatching OR trials, both visual and word-order cues support an incorrect reading. This pattern should be indexed by an interaction of sentence structure × visual cue (in the direction of larger visual-cue differences in response to OR vs. SR clauses). This finding would align with previous evidence from the dual-route framework (e.g., Ferreira, 2003) and extend it by including visual information as a relevant non-syntactic cue.

Lastly, the SSH predicts L2 speakers' reliance on visual cues to be at least comparable to L1 speakers', whilst L1 speakers should be more adept at using case cues during comprehension than L2 speakers. As in Experiment 1, this expected group difference should be especially pronounced in response to OR clauses, as responding accurately to canonical (SR) sentences can be achieved either via deep syntactic parsing or superficial com-prehension heuristics. In contrast, relying on an agent-before-patient strategy will lead to incorrect thematic-role assignment in noncanonical sentences.

4.7.1.2   Research question 2: The effect of morphosyntactic sensitivity on visual-cue weighting

Based on evidence for the role of individual-level profiles underlying cue weighting in L1 and L2 sentence processing (Hopp, 2015; Osterhout, 1997; Pakulak and Neville, 2010), we assume that individuals vary in their cue-weighting preferences even at the highest levels of proficiency (including L1 speakers). We hypothesize that the weighting of morphosyntactic and visual cues might be negatively correlated. Thus, participants who are more sensitive to case cues may be less sensitive to visual cues, and participants who are less sensitive to case cues may be more sensitive to visual cues. If so, increasing sensitivity scores should be associated with smaller differences between matching and mismatching trials, and vice versa.

4.7.2   Results

4.7.2.1   Descriptives

Average proportions of accuracy and RTs to correctly answered trials are reported per group and condition in Table 5. Descriptively, accuracy was relatively high for SR clauses across groups and visual-cue levels. Nonetheless, mismatching scenes caused accuracy rates to drop about 20% even in response to SR clauses in the L1 group, a difference that was smaller in the L2 group (~12%). In the L1 group, OR (vs. SR) clauses were associated with a 20% drop in accuracy, even with matching scenes, and for L2 speakers, accuracy dropped as much as 30%. These patterns are similar to those seen in Experiment 1. In the L1 group, accuracy dropped to the same extent for OR (vs. SR) clauses as in Experiment 1 in the match condition, while L2 speakers' performance dropped more steeply in Experiment 1 (40%) than in Experiment 2's match condition. In mismatching OR conditions, L1 speakers performed at chance level, whilst L2 speakers scored as low as less than a third of the total. The numerical differences between matches and mismatches were larger in OR (vs. SR) trials, and in L2 compared to L1 speakers. A descriptive comparison of the mean accuracy rates across clause types in both Experiments reveals both increased accuracy in response to matching scenes, and decreased accuracy in the presence of mismatching scenes, as compared to the same sentences presented without accompanying visual scenes. RTs are reported for completeness, but due to the low number of trials that would have been included in the RT analyses, these were not performed on the data from Experiment 2.

Table 5: Average proportions of accuracy and RTs in milliseconds to correctly answered trials in Experiment 2 for L1 and L2 speakers.

Group	Visual Cue	Sentence Structure	Accuracy	SD—Accuracy	RTs	SD—RTs
L1	Match	SR	0.93	0.25	1,096	377
L1	Match	OR	0.74	0.44	1,121	396
L1	Mismatch	SR	0.75	0.43	1,263	531
L1	Mismatch	OR	0.50	0.50	1,307	470
L2	Match	SR	0.89	0.31	1,250	454
L2	Match	OR	0.59	0.49	1,411	508
L2	Mismatch	SR	0.77	0.42	1,370	498
L2	Mismatch	OR	0.30	0.46	1,234	521

4.7.2.2   Research question 1: Visual influences on L1 and L2 sentence comprehension

To answer Research Question 1, we examined the effects of sentence structure, visual cue, group, and their interactions, which were added as fixed factors in the logistic regression fitted to accuracy data. The model's random-effects structure included by-item and by-participant intercepts, as well as by-participant Sentence and visual-cue slope adjustments. The coefficients of the model are reported in Table 6.

Table 6: Mixed-effects regression coefficients from the logistic regression fitted to accuracy rates from Experiment 2 (RQ1). sentence structure, visual cue, and group were sum-coded with ±0.5 weights, comparing the effect of OR (0.5) versus SR clauses (−0.5), of mismatch (0.5) versus match (−0.5), and of belonging to the L2 (0.5) versus L1 (−0.5) group. The intercept represents the grand mean.

Predictor	Estimate	Std. Error	z Value	p Value
(Intercept)	1.16	0.14	8.57	<0.001
Sentence Structure	−2.08	0.25	−8.47	<0.001
Visual Cue	−1.41	0.23	−6.11	<0.001
Group	−0.65	0.25	−2.59	0.009
Sentence Structure × Visual Cue	−0.37	0.31	−1.18	0.238
Sentence Structure × Group	−0.82	0.14	−1.78	0.074
Visual Cue × Group	0.04	0.42	0.11	0.914
Sentence Structure × Visual Cue × Group	−0.93	0.58	−1.61	0.106

Formula: accuracy ~ sentence structure × visual cue × group + (sentence structure + visual cue | subject) + (1 | item).

As expected, we detected simple effects of sentence structure, visual cue, and group. In line with results from Experiment 1, SR clauses were answered more accurately than OR clauses, and belonging to the L2 (vs. L1) group was associated with lower accuracy rates. In line with our predictions, matching trials were answered more accurately than mismatching trials. Further, we found some indication that L2 (vs. L1) speakers' performance was lower especially in response to OR trials, but in contrast to Experiment 1, the sentence structure × group interaction only reached marginal significance. However, post hoc tests indicated that group differences were significant in response to OR (b = 1.06, SE = 0.33, z = 3.16, p = 0.002) but not SR clauses (b = 1.13, SE = 0.48, z = 2.34, p = 0.019). Finally, the numerical tendency for

L2 speakers to show selective accuracy decreases in mismatching OR trials did not reach significance at an alpha level of $p < 0.05$.

4.7.2.3 *Research question 2: The role of morphosyntactic sensitivity for visual-cue weighting*

To answer Research Question 2, we fitted the same model as to answer Question 1, but we replaced L1–L2 group with the sensitivity scores obtained from Experiment 1. The model's coefficients are reported in Table 7.

Table 7: Mixed-effects regression coefficients from the logistic regression fitted to accuracy rates from Experiment 2 (RQ2). Sentence structure and visual cue were sum-coded with ±0.5 weights, comparing the effect of OR (0.5) versus SR clauses (−0.5) and of mismatch (0.5) versus match (−0.5). sensitivity scores were centered. The intercept represents the grand mean.

Predictor	Estimate	Std. Error	z Value	p Value
(Intercept)	1.20	0.11	10.52	<0.001
Sentence Structure	−2.06	0.25	−8.32	<0.001
Visual Cue	−1.44	0.23	−6.23	<0.001
Sensitivity score	4.65	0.64	7.21	<0.001
Sentence Structure × Visual Cue	−0.24	0.32	−0.76	0.449
Sentence Structure × Sensitivity Score	2.94	1.45	2.03	0.042
Visual Cue × Sensitivity Score	1.41	1.34	1.05	0.294
Sentence Structure × Visual Cue × Sensitivity Score	3.74	1.85	2.02	0.043

Formula: accuracy ~ sentence structure × visual cue × sensitivity score + (sentence structure + visual cue | subject) + (1 | item).

In addition to similar effects of sentence structure and visual cue as in the previous analysis, higher sensitivity scores were associated with higher accuracy overall, indicating that individual accuracy in Experiment 1 predicted accuracy in Experiment 2. Further, the interactive effects with group that were marginal or non-significant in the previous analysis reached significance as a function of sensitivity score: increasing sensitivity scores were associated with smaller sentence-structure differences in accuracy. This finding confirms our assumption that high levels of accuracy in Experiment 1 were associated with better ability to make use of subtle morphosyntactic cues for interpretation in the face of noncanonical structures. Moreover, sensitivity scores were significantly modulating match–mismatch differences depending on sentence structure (Figure 3). Post hoc tests showed that match–mismatch differences were greatest when participants at the lower end of the sensitivity range responded to OR clauses (b = 2.62, SE = 0.56, z = 4.67, p < 0.001). match–mismatch differences were also significant, yet less pronounced, when participants at either end of the sensitivity range responded to SR clauses (lower end: b = 1.17, SE = 0.58, z = 2.02, p = 0.043; upper end: b = 1.47, SE = 0.69, z = 2.12, p = 0.034). In contrast, match–mismatch differences were not significant when participants at the upper end of the sensitivity range responded to OR clauses (b = 0.50, SE = 0.55, z = 0.91, p = 0.362). We discuss this result in detail in the following section.
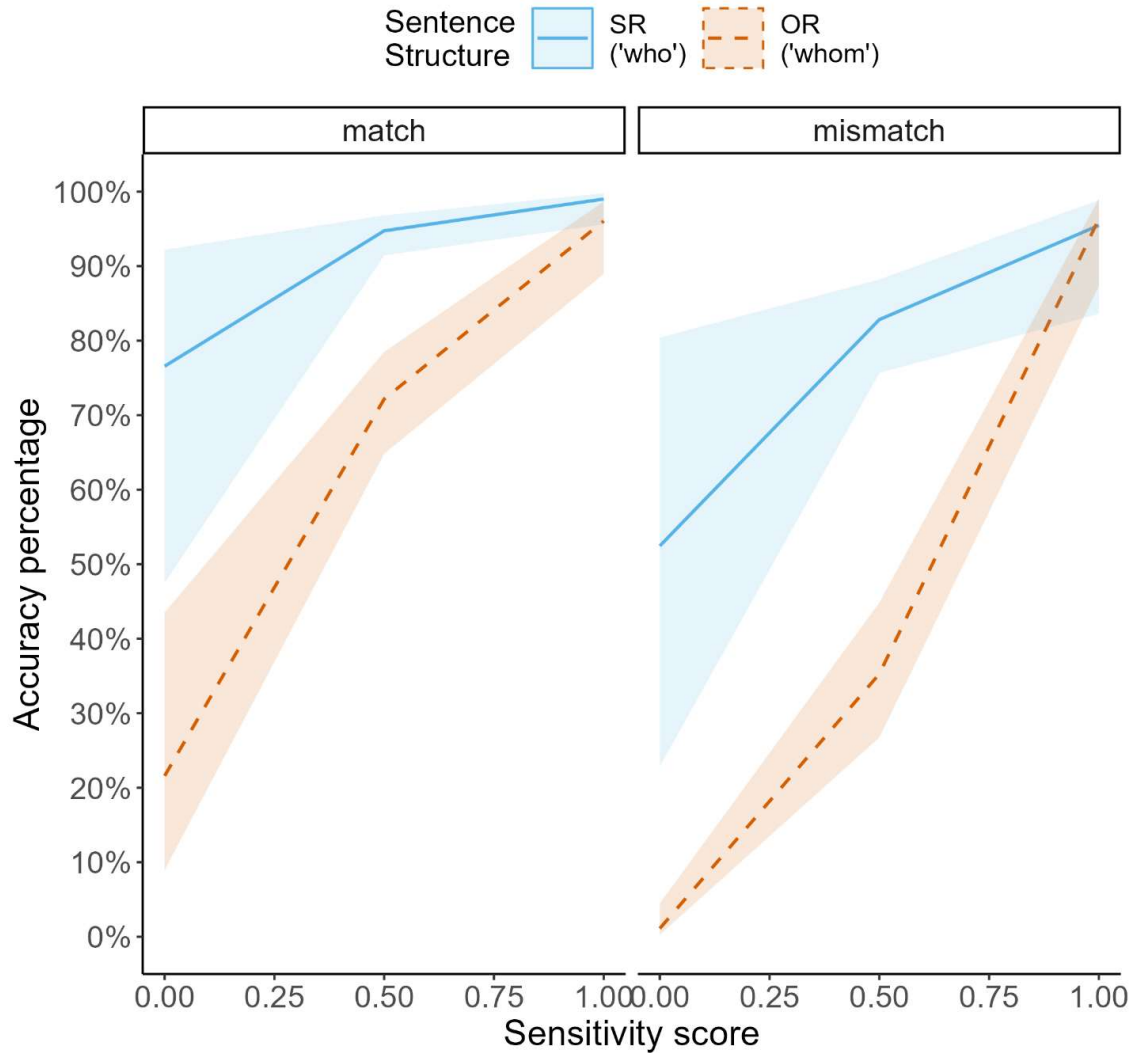
Fig 3 Interaction of sentence structure × visual cue × sensitivity scores across speakers. Error bands represent 95% confidence intervals

Finally, we contrasted the model including group to the model including sensitivity score with an ANOVA model comparison. The Akaike Information Criterion (AIC) of the 'sensitivity-score' model was lower (1765.7) than the AIC of the 'group' model (1805.1). This result confirms that sensitivity scores had more explanatory power as a predictor of comprehension differences in our dataset than L1–L2 grouping.

To summarize, we found that both our sentence structure and visual cue manipulations influenced performance, yielding higher accuracy for SR (vs. OR) clauses and for matching (vs. mismatching) visual cues. L1 speakers were more accurate than L2 speakers overall in both experiments, a pattern that was more pronounced in response to OR clauses (significant in Experiment 1, marginal in Experiment 2). Across groups, sensitivity scores modulated reliance on visual cues as a function of sentence structure. Low morphosyntactic sensitivity was associated with larger match–mismatch differences, particularly in response to OR clauses, while high sensitivity was associated with larger match–mismatch differences in response to SR than OR clauses. Finally, sensitivity scores were found to be a better predictor than group in modeling agent-identification accuracy.

4.8   Discussion

The present study investigated the processing of visual and morphosyntactic cues for thematic-role assignment of unambiguous German sentences in L1 and L2 comprehension. Further, we took an individual-difference approach to investigating the relationship between morphosyntactic sensitivity and the use of visual cues across L1 and L2 speakers. We asked (i) how (task-unrelated) visual cues influence thematic-role assignment of unambiguous sentences in L1 and L2 speakers, and (ii) whether a proxy of individual sensitivity to morphosyntax predicts sensitivity to visual information.

The novel findings from the current study (discussed in detail below) can be summarized as follows: first, visual-scene cues influenced agent-identification accuracy of L1 and L2 speakers alike, even though case marking unambiguously indicated thematic relations, and although the task was unrelated to sentence–picture congruence. Second, morpho-syntactic sensitivity modulated reliance on visual-scene cues as a function of sentence structure: Individuals with low morphosyntactic sensitivity tended to rely on visual-scene cues for thematic-role assignment in noncanonically ordered (and to a lesser extent, canonically ordered) relative clauses. In contrast, individuals with high morphosyntactic sensitivity relied on visual-scene cues for thematic-role assignment of canonical structures only. Third, we found that morphosyntactic sensitivity scores were a better predictor of agent-identification accuracy than categorical L1–L2 grouping.

4.8.1   Replication of subject preference and L1 advantage

In both experiments, noncanonical OR structures yielded lower accuracy (and in Experiment 1, slower RTs) than canonical SR structures. Our results, showing relatively high error rates in L1 speakers irrespective of syntactic ambiguity, align with previous evidence for accuracy decreases in the face of noncanonical but unambiguous sentences, particularly as measured by agent-identification tasks (Chromý, 2022; Ferreira, 2003). In the dual-route framework, accuracy decreases in response to unambiguous noncanonical structures are

attributed to reliance on 'fast-and-frugal' comprehension heuristics (e.g., word-order, frequency) that strongly favor interpretations that correspond to the 'agent-before-patient' pattern. In multiple-constraint models, the same basic finding is interpreted in terms of relative cue weightings. Crucially, replicating previous evidence for decreased agent-identification accuracy in response to noncanonical sentences in German (e.g., Bader and Meng, 2018, 2023; Meng and Bader, 2021) enables us to use the phenomenon as a testbed for assessing to what degree the reliance on morphosyntactic versus word-order cues to interpretation predicts reliance on visual cues.

In Experiment 1, L1 speakers were faster and more accurate at responding than L2 speakers, replicating well-established findings in L2 research. Further, we found more pronounced L1–L2 differences in accuracy to OR (vs. SR) clauses, in line with previous findings of disproportional difficulties triggered by noncanonical structures in L2 (vs. L1) comprehension (e.g., Gerth et al., 2017; Jackson, 2008; Pozzan and Trueswell, 2016 - but cf. Cunnings and Fujita, 2021). The latter interaction was only significant in the absence of visual cues, only reaching marginal significance when visual-scene cues were available. However, the group and sentence structure factors were involved in a relatively strong, higher-level numerical tendency including the visual-cue factor in Experiment 2. The discrepancy between the sentence structure × group interactions in Experiments 1 and 2 may be due to the modulation of co-occurring visual-scene information in Experiment 2. This conclusion is supported by the significant interaction obtained as a function of sentence structure, visual cue, and sensitivity scores in Experiment 2, reflected by a large numerical tendency when group was included as predictor instead of sensitivity (discussed below).

In line with the SSH, our finding of accuracy decreases in the L2 group's responses to OR clauses can be explained in terms of reduced sensitivity to case-marking cues, and correspondingly greater reliance on linear word-order, in L2 vs. L1 comprehension. In response to canonical SR clauses, we found no difference between L1 and L2 speakers' performance, indicating that comprehension should be unproblematic for L2 speakers when word-order and case cues converge. In contrast, selective difficulties in L2 comprehension of noncanonical structures may be ascribed to L2 speakers' enhanced reliance on shallower, word-order-based heuristics.

4.8.2   Research question 1: Visual influences on L1 and L2 sentence comprehension

In line with predictions from the CIA, mismatching (vs. matching) visual scenes yielded lower agent-identification accuracy in both L1 and L2 speakers. This confirms that visual cues are relied upon in both the L1 and L2 comprehension of unambiguous structures. As mentioned above, this finding complements and extends existing evidence on the integration of visual and morphosyntactic cues. First, the present study assessed the influence of visual

cues on final interpretations, whereas previous studies using unambiguous linguistic stimuli assessed the time course of visuo-linguistic integration during processing using the VWP (e.g., Knoeferle and Crocker, 2006). Contrary to the VWP, however, our setting enabled us to present visual-scene cues that were incompatible with the action direction conveyed linguistically. We further measured comprehension using a linguistic task unrelated to sentence–picture congruence, unlike studies using sentence–picture verification (e.g., Underwood et al., 2004). Though Knoeferle and Crocker (2006) have provided evidence for a role of sentence–picture congruence using reading times as a purely linguistic outcome measure, their study differed from the current investigation in that they presented ambiguous sentences in written form.

In short, the present study's design choices were justified by the attempt to provide evidence for the effects of visual cues on sentence comprehension in a 'maximally conservative' setting. In other words, we strove to provide a potentially unfertile ground for visual influences on comprehension to emerge. The fact that we nevertheless found reliable effects of visual-scene cues in the present setting provides strong support for the robustness of visual influences on sentence processing. Descriptively, we found that matching scenes increased, whilst mismatching scenes decreased, comprehension accuracy relative to conditions in which no accompanying visual information was presented. This observation supports the view that matching visual input can be advantageous for comprehension. It is possible that matching scenes allowed participants to form richer mental representations as compared to language alone, thus improving post-sentence recall of thematic-role relations. In contrast, mismatching scenes should have triggered an attempt to re-analyze (CCM) or reconcile (CIA) spoken and visual input, which we found to negatively influence post-sentence recall of the role-relations in the sentence. Thus, the abstract relations depicted in the visual scenes were integrated in the computation of meaning representations, irrespective of task requirements or syntactic ambiguity.

The effect of visual cue was not significantly modulated by sentence structure. Rather, morphosyntactic and visual cues seemed to contribute additively to agent-identification accuracy. This result contrasts with evidence in favor of the influence of visual cues (e.g., speaker gaze, visual scenes) on the comprehension and processing of noncanonical structures (e.g., Knoeferle and Kreysa, 2012; Kreysa et al., 2018). In this research, matching visual cues were found to neutralize the preference for canonical over noncanonical syn-tax in anticipating upcoming referents, which was not the case in our results. Instead, the additive patterns in our results seem to support the view that linguistically and visually derived representations might be stored in a modality-dependent fashion, rather than in a common representational format. However, this null result is only suggestive, and conclusive insights in this regard would have to be based on positive evidence. Further, the lack of an interaction of visual cue and sentence structure contrasts with dual-route claims about increased reliance

on non-syntactic cues in the face of noncanonical syntax in L1 processing (Christianson et al., 2010; Ferreira, 2003).

To our knowledge, the current study represents a novel attempt to explore the applicability of visuo-linguistic integration models (such as the CIA) to L2 sentence comprehension. Our results therefore add to the generalizability of CIA claims across populations. Interestingly, we found that the size of the visual effects overall was statistically in-distinguishable between the L1 and L2 groups. This indicates that L2 speakers were as able as L1 speakers to utilize visual cues to interpretation.

4.8.3   Research question 2: Role of morphosyntactic sensitivity for visual-cue weighting

We found that comprehension accuracy in Experiment 1 (used as a proxy of morphosyntactic sensitivity) was also associated with higher accuracy in Experiment 2, both overall and specifically in response to noncanonical sentences. First, participants who re-lied on case cues in Experiment 1 also relied on this type of interpretation cue in Experiment 2. Second, the selective effect of sensitivity score on comprehension of OR sentences suggests that surface-level heuristics were underlying thematic-role mis-assignment. This is because SR clauses present no conflict between the agent-before-patient preference on the one hand, and case cues on the other. Thus, SR-OR differences were minimal in participants who were able to use case cues for interpretation (i.e., participants with a high sensitivity score). Lastly, the finding that sensitivity to morphosyntactic cues explains L1–L2 variability better than L1–L2 grouping indicates that L1–L2 processing differences are not necessarily categorical but may be gradient (compare, e.g., Clahsen and Felser, 2018).

We now turn to the observed three-way interaction of sentence structure × visual cue × sensitivity scores. This interaction confirms our hypotheses as to the negative association of morphosyntactic sensitivity and reliance on visual-scene information: the less participants were able to use morphosyntactic information, the more they relied on visual-scene information for comprehension. The interaction reflects the following: While the visual-cue effect size was similar in high and low-sensitivity participants in SR trials, in OR trials the visual-cue effect was absent in high-sensitivity participants. In low-sensitivity participants, in contrast, the visual-cue effect size almost doubled (as compared to SR trials). These patterns broadly confirm our prediction of a negative association between sensitivity to case and visual cues, thus adding to current efforts to understand the role of individual variability in L1 and L2 sentence comprehension. This evidence aligns with the results from Hopp (2015), where speakers' ability to integrate syntactic versus semantic cues for sentence processing was found to correlate negatively at the individual level. In the pre-sent study, speakers' sensitivity scores were found to determine comprehension success across both L1 and L2 speakers, as

reflected by the disproportional influence of visual cues on the comprehension of noncanonical clauses at lower levels of morphosyntactic sensitivity.

Our finding that highly sensitive participants showed greater visual-cue effects in SR than OR trials, however, seems counterintuitive. One possibility is that processing OR clauses, where the agent-before-patient preference conflicts with case cues, takes up more cognitive resources than processing SR trials, where there is no such conflict. This would then leave them with fewer cognitive resources in OR trials for taking into account visual context information, in comparison to SR trials and to participants with lower morpho-syntactic sensitivity.

To summarize, our results confirm the claims made by the CIA and extend their generalizability by testing them in a strictly conservative experimental setting, as well as providing evidence for their validity from a novel population. Further, our results support our hypothesis as to a trade-off in the reliance on morphosyntactic and visual cues across speakers. Our results demonstrate that differences in successful thematic-role assignment can be better accounted for by individual differences in cue weighting (Clahsen and Felser, 2018; Cunnings, 2017) than in terms of L1–L2 grouping.

4.9   Conclusions

Our findings expand on previous evidence for the influence of visual cues on the comprehension of syntactically unambiguous sentences in L1 and L2 speakers, emphasizing the need for the integration of visual cues in research on language comprehension. Our results further support the view that differences in the weighting of syntactic and non-syntactic cues underlie L1–L2 group differences, whilst highlighting the usefulness of determining individual cue-weighting profiles for explaining within-group differences in sentence comprehension. Future research should address individual differences within L2 speakers with larger samples and seek to identify potential cognitive or environmental factors underlying cue-weighting profiles in L1 and L2 comprehension.

5   Publication 3

Under review in *Psychology and Aging*.

The weighting of syntactic versus visual-context information during comprehension: A lifespan perspective

Carlotta Isabella Zona[1] and Jana Reifegerste[2]

(1)   Potsdam Research Institute for Multilingualism, University of Potsdam, Germany
(2)   Department of Neurology, Georgetown University, Washington DC, USA

5.1   Abstract

Current models of visuo-linguistic integration (Knoeferle & Crocker, 2006, *Cog Sci*) predict that young adults attempt to reconcile visual and syntactic cues during sentence comprehension. Yet, it remains to be elucidated how such integration plays out in older adults, who might weight non-syntactic (e.g., semantic) information more heavily (Beese et al., 2019, *Psych Aging*). To investigate whether older adults rely more on visual (as on semantic) cues, we examined the role of scene depictions for thematic-role assignment of unambiguous sentences across the adult lifespan. 154 German speakers (18-75 years) listened to spoken subject- and object-relative clauses ("…the man who is following the woman"/"the man whom the woman is following") presented in isolation (baseline) or alongside depictions conveying the same (matching) or the opposite action direction (mismatching) as the sentence. After each sentence, participants responded to binary agent-identification questions. Overall, matching scenes yielded greater accuracy, while mismatching scenes yielded accuracy declines relative to baseline. The size of the mismatch effect increased with increasing age. Increasing age was also associated with smaller accuracy differences between subject- and object-extracted clauses, and steeper RT increases in baseline trials than when presented alongside visual scenes – possibly due to lexical facilitation. Our results demonstrate that visually encoded thematic relations influence thematic-role assignment of unambiguous syntax, and that such effects may become more pronounced with increasing age. We argue for age-related increases in the integration of cross-modal cues for sentence comprehension (rather than processing limitations) as a potential cause for such effects.

5.2   Introduction

Every day, we are surrounded by visual stimuli alongside linguistically encoded information in spoken or written form. Most, if not all, linguistic processes are situated in an environment with a visual component: a conversation in a crowded café, watching a video with voice-over, or reading a newspaper on the subway – all of these are highly cross-modal experiences, during which information is encoded both visually and linguistically. While in some situations, the use of visual information might facilitate the understanding of linguistically encoded meaning (e.g., to resolve ambiguities or to follow instructions), in many scenarios visual information might not be important for the processing of linguistically encoded information, and perhaps even at odds with it. The present study seeks to elucidate the latter situation, in which visual information is irrelevant for the task at hand, and moreover incompatible with linguistically conveyed meaning.

While most prior psycholinguistic research has examined language in isolation, there has been increasing consideration of language as immersed (or 'situated') within a perceptual environment. In particular, visual cues appear to guide the anticipatory processing of lexical

and syntactic information and may affect sentence comprehension and recall (see Knoeferle & Guerra, 2016; Knoeferle, 2019, for review and discussion). However, while the interaction between language and visual cues during early stages of development has been studied extensively (Borovsky et al., 2012; Eiteljoerge et al., 2019; Tomasello et al., 2007; Wellsby & Pexman, 2014), only a handful of studies have examined 'visually situated language' in aging populations (Münster & Knoeferle, 2017).

There appears to be a tacit assumption that aging is associated predominantly with cognitive declines; however, recent research has painted a finer-grained picture. Though senescence drives declines in various aspects of cognition, most notably processing speed (Shake & Stine-Morrow, 2016) and episodic memory (see, e.g., Pelletier et al., 2017; Reifegerste, Veríssimo et al., 2021), compensatory brain mechanisms (Thomas et al., 2019) and an ever-growing knowledge- and experience-based competence can allow for the stable maintenance and even improvements of some aspects of cognition (Nyberg et al., 2012; Rey-Mermet & Gade, 2018; Sánchez-Izquierdo & Fernández-Ballesteros, 2021; Veríssimo et al., 2022). Research on language processing across the lifespan has sought to identify which aspects of language may change during aging, and whether they show improvements or declines, as well as the cognitive skills underlying specific linguistic processes (whose respective developmental trajectories may indeed moderate or mediate age-related changes to language processing).

Typically, lexical-retrieval difficulties are considered a hallmark of age-related declines (see, e.g., Connor et al., 2004; Verhaegen & Poncelet, 2013), while the evidence for other language domains, such as syntax, is more mixed (Campbell et al., 2016; Hardy et al., 2020; Poulisse et al., 2019), with some studies reporting similar performance between younger and older adults, while others report age-related declines in syntactic processing abilities. Of relevance for the current study, it appears that when non-syntactic cues (e.g., plausibility) are available and they support the same interpretation of a syntactic construction, older adults may respond to, for example, morphosyntactic cues similarly to younger adults (Amichetti et al., 2016; Caplan & Waters, 2005; Campbell et al., 2016), suggesting that access to information from non-syntactic sources may be relatively stable throughout the lifespan, even in the face of possible age-related declines in syntactic-processing abilities. One explanation that has been suggested to account for such performance stability across the adult lifespan is age-related shifts in the information sources (or 'heuristics') that older adults rely on during sentence comprehension (Beese et al., 2019; van Boxtel & Lawyer, 2021). The current study aims to explore whether non-syntactic cues that are not linguistic in nature, such as visual cues, are increasingly relied on by older (vs. younger) adults, and how visual and linguistic cues interact in guiding comprehension. To this aim, we examined how scene depictions may affect thematic-role assignment of globally unambiguous relative clauses in speakers from across the adult lifespan.

In the remainder of the Introduction, we first review the "Good-Enough" approach to sentence comprehension, which has examined the (mis)interpretation of non-canonically ordered, yet unambiguous syntactic structures (section 'The role of non-structural cues'), which provides an apt framework guiding the hypotheses for the present study. Second, we summarize findings on the use of visual information in sentence comprehension, with a focus on lifespan trajectories (section 'The role of visual cues').

5.3   Background

5.3.1   The role of non-structural cues: the "Good-Enough" approach

The Good-Enough approach posits that comprehenders may form sentence representations that are inconsistent with the linguistic input, and especially when parsing noncanonical syntactic structures (Ferreira & Patson, 2007; Sanford & Sturt 2002). For instance, Ferreira (2003) examined the comprehension of unambiguous subject and object clefts in English (e.g., object cleft: *It was the man the dog bit*), while manipulating pragmatic plausibility (e.g., implausible: *It was the dog the man bit*). Agent-identification accuracy was higher for subject clefts overall, and the effect of plausibility (greater accuracy for plausible vs. implausible sentences) was evident in responses to object (but not subject) clefts. These results suggest that the meaning of noncanonical sentences may be inferred from linear word order, and that lexico-semantic cues supporting a specific reading may play a role in modulating which interpretation is endorsed (Karimi & Ferreira, 2016). Findings in this domain are often interpreted as evidence for a dual-processing strategy of sentence comprehension (e.g., van Boxtel & Lawyer, 2021; Kuperberg, 2007). One route, syntactic analysis, derives language meaning from (morpho)syntactic information, such as subject-verb agreement and case marking, among others. The second route, instead, makes use of word-order and lexico-semantic cues, which is thought to be less demanding than a full syntactic analysis.

While word order is the strongest cue for thematic-role assignment in English (MacWhinney et al., 1984), German additionally uses morphological case markings on pronouns and noun-phrase (NP) determiners (and to some extent via noun suffixes). These may allow comprehenders to unambiguously assign thematic roles to constituents, resulting in more flexible word order in German than English: While noun-verb-noun structures can often be correctly mapped onto SVO structures in English, they may either correspond to SVO or OVS structures in German. Thus, morphological case cues unambiguously assigning thematic roles represent an important test case to investigate the influence of non-syntactic information on thematic-role assignment, as they may neutralize the effect of biasing non-syntactic information for comprehension (Schlesewsky & Bornkessel, 2004). However, contrary to this prediction, Meng & Bader (2021) reported significantly decreased agent-identification accuracy in response to unambiguous German OS (vs. SO) sentences. These effects were present whether the thematic roles were reversible (e.g., 'The father hugs the

uncle'), biased (e.g., 'The chef ruined the roast'), or non-reversible (e.g., 'The chef cleaned the pan'). Closely replicating previous research (Bader & Meng, 2018; Ferreira, 2003), these results suggested a crucial role of linear word order for agent-identification success and showed that unambiguous case markings may not fully override the 'agent-before-patient' preference in German thematic-role assignment.

Findings regarding sentence processing in aging are somewhat mixed. On the one hand, studies tapping different syntactic phenomena have found age-related performance declines in sentence comprehension (e.g., Christianson et al., 2006; Poulisse et al., 2019), particularly in conditions of increased task demands (Beese et al., 2019a; Reifegerste & Felser, 2017) or for participants with relatively decreased working memory (WM) capacity (Peelle et al., 2010; Stine-Morrow et al., 2000; Reifegerste et al., 2017). On the other hand, no age changes were reported in the ability (i) to revise interpretation preferences based on agreement (Samu et al., 2017), as well as (ii) "to develop syntactically and semantically coherent sentential representations" (Tyler et al., 2010, p. 356; Davis et al., 2014). For instance, Wingfield et al. (2003) found no age-related differences in the comprehension of subject- vs. object-extracted clauses.

Further, some evidence has pointed to lifespan changes in comprehenders' reliance on cues from different sources for text and sentence comprehension (Kemtes & Kemper, 1997; Stine-Morrow et al., 1996). Older (vs. younger) adults have shown equal or increased likelihood to rely on non-syntactic cues during comprehension (e.g., Amichetti et al., 2016, Exp. 1; Ayasse et al., 2021; Beese et al., 2019; Christianson et al., 2006; Poulisse et al., 2019; Rangel Ferrari et al., 2019; Yoon et al., 2015). For instance, Beese et al. (2019; Exp. 2) compared younger and older adults' ability to use morphosyntactic and semantic constraints for sentence comprehension. Participants were presented with written sentences and word lists of either meaningful or non-words, and they answered questions probing serial order position (e.g., 'Did the word *X* come before the word *Y*?'). No age differences in performance were reported with the word lists, or in the extent to which meaningful words (vs. non-words) facilitated task performance. These findings indicate that the ability to use semantic cues for the task was unaffected by increasing age. Thus, it has been proposed that older (vs. younger) adults may experience a "syntax-to-semantics" shift, i.e., they may achieve comprehension by relying more on semantic relations between different sentence elements (Taler et al., 2009), plausibility (Amichetti et al., 2016), and other domains related to crystallized knowledge (Campbell et al., 2016). Against this background, the current study aims to explore whether visual input might be another non-syntactic information source that is relied on equally or more with increasing age.

To summarize, the findings from this research suggest that misinterpretations of noncanonical structures – that is, where linear word-order conveys a reading that is not licensed by the syntax – are pervasive, especially as measured by agent-identification tasks.

Such misinterpretations are more likely to occur when additional non-syntactic cues converge with linear word-order in conveying a reading not licensed by case cues. The current study examines whether visual information can be included (i) as one of the non-syntactic cues constraining sentence processing in the Good-Enough approach, and (ii) as a cue increasingly relied upon by older adults during sentence comprehension.

5.3.2   The role of visual cues: The Coordinated Interplay Account and Situation-Model theory

'Situated' language research has considered language as naturally immersed in a visual environment, highlighting the reciprocal influence of visual and linguistic processing, and suggesting a close temporal link of utterance comprehension and visual-attention allocation (see Knoeferle, 2019, for a review). To account for this evidence, the Coordinated Interplay Account (CIA) proposed a cyclic mechanism in two stages underlying visuo-linguistic integration (Knoeferle & Crocker, 2006). First, as utterances unfold, comprehenders direct their attention towards the relevant referents in the scene. Second, the relevant features of the visual scene (e.g., the action direction) are integrated with the linguistic input and rapidly influence its processing.

While a few studies have examined the lifespan trajectory of the role of visual information for sentence comprehension within the CIA (for review, see Münster and Knoeferle, 2017), Situation-Model theory has considered age-related changes in the use of perceptual features more systematically (Radvansky, 1998; Zwaan and Radvansky, 1998). According to Situation-Model theory, the "syntax-to-semantics" shift may be explained in terms of older adults' ability to retrieve information from *situation models*, i.e., from "the representations of situations as they would exist in a real or possible world. As such, [situation models] include […] inferences generated from world knowledge" (p. 1). Crucially, older adults have been shown equal or better memory performance for information extracted from situation models relative to younger adults (Radvansky et al., 2001, 2003). Some research has investigated age-related changes in the reliance on visual cues in building situation models.

Dijkstra et al. (2004) presented written sentences to 19 younger and 19 older participants (e.g., *There is an eagle [in the sky/on the tree]*). After each sentence, participants were shown a picture of an object that either matched or mismatched the shape of the object implied by the sentence (e.g., a picture of an eagle depicted either as in flight or perched) and were asked to quickly judge whether the object had been mentioned in the sentence. When the shape of the object implied in the sentence and the depicted object mismatched, older participants showed a disproportionate reaction time (RT) increase as compared to younger participants. This effect suggests that older adults relied on visual cues to a larger extent than younger participants did, engendering increased difficulty when visual and linguistic cues mismatched. This finding supports the CIA, in that visual and linguistic representations are

checked against one another, and it suggests that this process is not only maintained at older age but might in fact occur to an increased extent. Similarly, Madden & Dijkstra (2010) asked participants to listen to object descriptions, then see object depictions (which were either congruent or incongruent with the description) and name the objects. Older (vs. younger) adults experienced greater benefits from congruent (vs. incongruent) shape features. Together, this evidence indicates that the ability to build and exploit mental situation models to support comprehension is stable, and may even improve, throughout the adult lifespan. Further, it suggests that older (vs. younger) adults might show increasing reliance not only on semantics, but also on visual stimuli.

Finally, Maguinness et al. (2011) directly contrasted younger and older adults' reliance on semantic and visual information. The authors presented 20 younger and 21 older participants with meaningful and non-meaningful spoken sentences (e.g., #*She packed her skyscraper for her trip*) as they watched video recordings, which were either blurred or clear, of lip movements corresponding to the spoken sentence. Participants were then asked to judge whether the sentence was meaningful, and to repeat the sentence aloud. Older (vs. younger) adults showed lower accuracy at detecting non-meaningful sentences only. Further, clear videos yielded better performance in the sentence-recall task for the older group only, and only in their recall of non-meaningful sentences. The authors concluded that with increasing age, semantic and visual cues played an increasing role in supporting sentence processing and recall: when the semantic component was lacking (i.e., in non-meaningful trials), age differences emerged, and these were further exacerbated in the absence of reliable visual cues.

To summarize, evidence on the integration of conflicting visual and linguistic cues for sentence-related processes across the lifespan is still limited and based on relatively small-scale studies. What is more, previous studies have typically operationalized age as a categorical variable. The present study seeks to address these gaps by investigating the lifespan changes in the use of visual and morphosyntactic cues for comprehension with a larger sample size than previous studies and treating age as a continuous variable.

5.4   Methods

5.4.1   Design

Our study aims to answer the following research questions:

1.   Do visual cues influence reliance on word-order versus morphosyntactic cues for the comprehension of grammatically unambiguous sentences?

2.   Are there age-related differences in the degree to which visual cues influence reliance on word-order versus morphosyntactic cues for comprehension?

To answer these questions, we tested a cross-sectional sample of participants from across the adult age range, ranging between 18 and 80 years of age. Using a 2 x 3 design, we manipulated the syntactic structure of spoken relative clauses in German (2 levels: subject- vs. object-extracted; henceforth, SR vs. OR), which were presented in isolation ('Baseline' condition; presented in a separate block) or alongside scene depictions that either matched or mismatched the action direction conveyed by the sentence (2 levels: Match, Mismatch; presented in the same block). After each sentence, we collected accuracy and RTs of participants' responses to a sentence-final written forced-choice comprehension question targeting the thematic roles conveyed in the sentence participants had heard.

5.4.2   Participants

Participants were recruited via the online participant recruitment platform of the University of Potsdam, via web-based ads throughout Germany, and through word of mouth. Participants provided informed consent to participate in the study and received 10€/h or university credit as compensation. Participants were screened for inclusion in the study with the German version of the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005; German version: Thomann et al., 2018) to limit the inclusion of participants with signs of age-related neurological impairment.

The final dataset included 155 participants, 100 women and 55 men. Figure 1 illustrates the distribution of participants by gender over the age range (18-80). All participants were native speakers of German and had not learned another language before the age of 5. Participants had received formal education for an average of 16 years (range: 7-25) and reported no perceptual, cognitive, or language-related impairments, and normal or corrected-to-normal vision (including intact color vision). A detailed breakdown of the participant information and exclusion procedures are available on OSF (https://osf.io/z9ach/?view_only=4b8e8ca7fa404a8d9e48288977289aac).
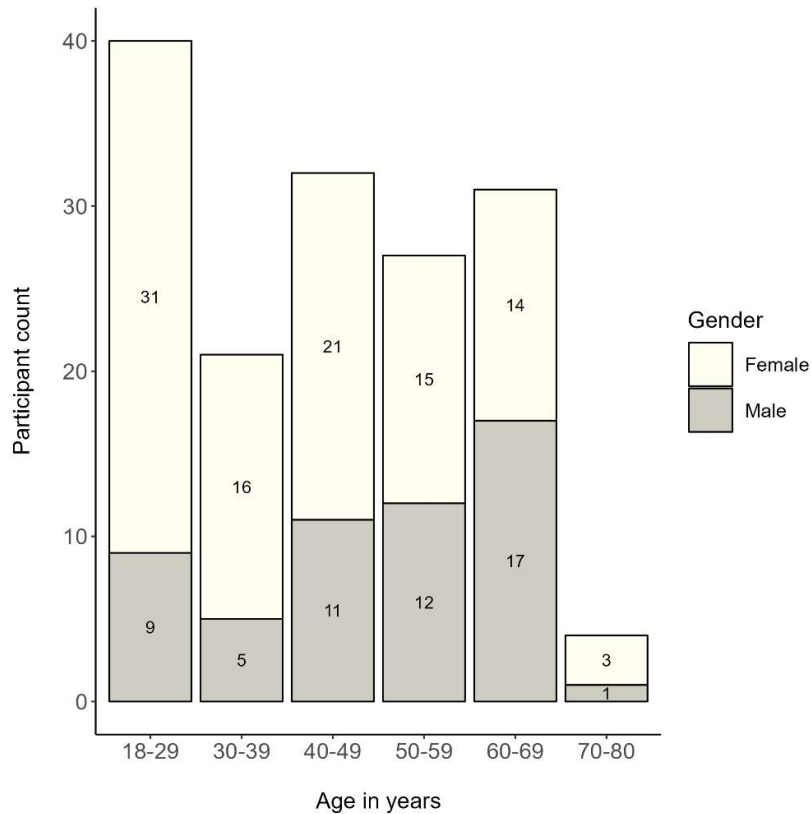
Fig 1 Participant count by age (in years) and gender. Note that participants are binned into decade bins for exposition only; all analyses treat Age as a continuous variable

5.4.3   Materials

The task was administered to each participant in two versions, once in the 'baseline block' and once in the 'critical block.' In the baseline block, participants saw a fixation cross while listening to spoken stimuli, whereas in the critical block, a single (matching/mismatch) scene depiction was shown on-screen as each spoken sentence was being presented. Materials and presentation lists are available on OSF (https://osf.io/z9ach/?view_only=4b8e8ca7fa404a8d9e48288977289aac).

The linguistic stimuli included 80 spoken sentences recorded by a female native speaker of German. Participants heard the same filler items in both blocks, while the target sentences were presented as SR in one block and as OR in the other. Each presentation list contained eight practice items, 24 target items, 16 pseudo-fillers, and 32 fillers. Item order within each list was fully randomized for each participant.

Target sentences were copular sentences with a defining relative clause. The main clause introduced a human male referent, followed by a relative clause introducing a (human) female referent and a transitive verb. See Table 1 for examples ('Auditory stimulus'). All referents (in the main clause and the relative clause) were common nouns indicating a profession (e.g., *der Koch* 'the cook'), a rank of nobility (e.g., *die Prinzessin* 'the princess'), or otherwise

stereotypically gendered terms (e.g., *die Braut* 'the bride'), and had masculine or feminine gender. For the manipulation of syntactic structure, half of the relative pronouns were in nominative case (*der* 'who') and half in accusative case (*den* 'whom'), thereby forming subject-extracted (SOV) and object-extracted (OSV) relative clauses, respectively. Only masculine nouns were included as pronoun antecedents because masculine singular pronouns always allow for unambiguous thematic role assignment in German, while feminine and neuter relative pronouns as well as plurals render sentences ambiguous due to case syncretism. Filler (n=32) and pseudo-filler items (n=16) were added to prevent participants from developing response strategies, and included human, animal, and inanimate referents, as well as visual mismatches of different types (e.g., action-verb, number mismatches). More details on the fillers are available on OSF ('Experiment Documentation').

Scene depictions consisted of photos of Playmobil© figurines performing actions, against a white background. The photos were taken in a brightly lit room with a Canon EOS 40D digital camera and edited with the open-source software GIMP (GNU Image Manipulation Program) software (version 2.10.20; The GIMP Development Team, 2019). All pictures were cropped to a 1:1 width-to-height ratio, and their visual features were kept as similar as possible across the different conditions (e.g., characters' size and location on the screen, their respective position, brightness, saturation of color features, etc.). Each target image depicted only the two referents engaging in the action described by the sentence. For target sentences, visual scenes in the Mismatch condition deviated from the meaning of the sentence only with regards to the conveyed role relations (i.e., the action direction). See Table 1 for examples of target pictorial stimuli ('Visual Cue') and the 'Experiment Documentation' (on OSF) for visual mismatches in (pseudo-)filler items.

Each sentence was directly followed by a slide with a written probe question (in passive voice) which targeted agent identification. Half of the probe questions required a *Yes* and half a *No* response. See Table 1 for example questions ('Written probe question'). Half of the questions referred to the male and half to the female character.

Table 1: Experimental design and examples.

First slide				
Sentence structure	Subject-relative clause (SR)		Object-relative clause (OR)	
Visual cue	Match	Mismatch	Match	Mismatch
Auditory stimulus	*Das ist der Koch, **der** die Braut verfolgt.*	*Das ist der Koch, **der** die Braut verfolgt.*	*Das ist der Koch, **den** die Braut verfolgt.*	*Das ist der Koch, **den** die Braut verfolgt.*
	This is the$_{NOM}$ cook, who$_{NOM}$ the$_{ACC}$ bride follows.	This is the$_{NOM}$ cook, who$_{NOM}$ the$_{ACC}$ bride follows.	This is the$_{NOM}$ cook, who$_{ACC}$ the$_{NOM}$ bride follows.	This is the$_{NOM}$ cook, who$_{ACC}$ the$_{NOM}$ bride follows.
	'This is the cook who is following the bride.'	'This is the cook who is following the bride.'	'This is the cook whom the bride is following.'	'This is the cook whom the bride is following.'
Visual cue				
Second slide				
Written probe question	*Wird die Braut/der Koch verfolgt?* Is the$_{NOM}$ bride/the$_{NOM}$ cook followed? 'Is the bride/the cook being followed?'			

5.4.4   Procedure

The experiment was programmed in E-Prime 3.0 (Psychology Software Tools, 2016) and conducted remotely during the COVID-19 pandemic using E-Prime Go 1.0 (Psychology Software Tools, 2020). All participants completed an online screening questionnaire (see section 'Participants'), and then received an e-mail with instructions, a download link for the experiment, and a subject ID number. An extensive instruction manual (available in German on OSF) was also provided to help participants navigate the experiment and trouble-shoot in case of technical problems. None of the tasks provided feedback to the participants.

Prior to the experiment, participants filled out a demographic background questionnaire, asking for their gender, date of birth, education (in years), native language, other languages spoken (specifying age-of-acquisition and acquisition context), handedness, vision/hearing deficits, and (past or current) language-related/cognitive/neurological impairments.

Task administration was arranged in two parts. The first part included the MoCA, a Stroop task (measuring interference inhibition; see below for details), and a 1-back task (measuring working-memory span; see below for details), presented to each participant in random order. The tasks were presented in black text against a white background in Consolas font size 18. (See 'Experiment Documentation' on OSF for more information https://osf.io/z9ach/?view_only=4b8e8ca7fa404a8d9e48288977289aac). The second part

included two sentence-comprehension tasks administered in separate blocks (baseline and visual-cue trials), presented in counterbalanced order. Participants were allowed to take a break between cognitive-task and sentence-comprehension parts, as well as between both sentence-comprehension sub-blocks.

Each trial in the sentence-comprehension tasks consisted of two steps (see Table 1: 'First slide,' 'Second slide'): on the first slide, participants heard a spoken sentence; the second slide presented a written binary forced-choice probe question that participants were asked to answer through keypress within 4000ms (left-hand side: *Yes*; right-hand side: *No*; reminders of which key represents which response stayed on-screen throughout the task). As mentioned above, the task was administered once in the baseline block (i.e., with no visual cues) and once in the critical block (i.e., with matching/mismatching visual cues), with the order of these two blocks counterbalanced between participants. Importantly, participants were asked to listen to the sentences and answer the questions in relation to the sentence only; they were not specifically instructed to attend or disregard the visual scenes. While target items were presented in a Latin-Square design, filler and pseudo-filler sentences were presented in the same version in the baseline block and the critical block, though the correct response to the comprehension question was counterbalanced between blocks (by changing the subject of the comprehension question) to prevent habituation. Participants completed eight practice trials before each of the two blocks; four of the practice trials resembled target and pseudo-filler items, and four resembled filler items. Practice trials were distributed symmetrically across conditions: half of target-like trials contained subject-relatives and half object-relative sentences; in addition, both filler- and target-type trials were presented alongside half matching, half mismatching depictions. Practice probe questions required an equal number of *Yes* and *No* responses. Items in each task were presented to each participant in fully randomized order. The total duration of both blocks was about 25 minutes.

5.4.5   *Data analyses*

All analyses were performed in R 2021.09.0 (Rstudio Team, 2022). The dependent measures were accuracy and RTs (for correct responses) in the sentence-comprehension tasks. We calculated mixed-effects logistic regression models (binomial family) for accuracy and linear mixed-effects regression models for RTs (natural-log-transformed). Only RTs to correctly answered trials were analysed. All regression analyses were carried out with the *lme4* R package (Bates et al., 2015), figures based on model estimates were plotted using *ggplot2* (Wickham, 2016).

For both accuracy and RT models, Sentence Structure (SR, OR), Visual Cue (ordered: Match, Baseline, Mismatch)[2], Age (in years; continuous), and their interactions were included in the models as fixed factors of interest. Sentence Structure was sum-coded with ±0.5 weights. Successive-difference contrast coding was used for the factor Visual Cue to allow for the comparison of the Match and the Mismatch condition to the same 'Baseline' level. Hence, in what follows, we distinguish these comparisons by referring to the effects of "Visual Cue (Match vs. Baseline)" and "Visual Cue (Mismatch vs. Baseline)," respectively. The choice of comparing Match and Mismatch conditions against a neutral Baseline allows us to assess the directionality of Visual-Cue effects (i.e., whether matches aided, and/or mismatches hindered, comprehension), as well as its lifespan trajectory. Task presentation order (i.e., whether participants were administered first the Baseline or the Visual-Cue block) was included as a fixed factor following a reviewer's suggestion and sum-coded with ±0.5 weights. See the captions of the results tables for further information on coding.

Follow-up analyses were performed with the *emmeans* package (Lenth, 2022). Following previous research (e.g., Ashaie & Obler, 2014; Connor et al., 2004; Veríssimo et al., 2022), we also tested for non-linear effects of Age. Since no significant quadratic or cubic effects of Age emerged, higher-order polynomials were not tested for inclusion.

5.5   Results

5.5.1   *Descriptives*

Figure 2 displays by-participant accuracy (in percentage correct; Panel A) and raw RTs (in milliseconds; Panel B) data, by age and by condition. The data are available in tabular form on OSF. Descriptively, increasing age seemed to be associated with increasing RTs, though the participants over the age of 70 seemed to respond more quickly. The age-related RT increases were comparable for both clause types, while accuracy rates seemed to change throughout the lifespan less homogeneously. Pearson's product-moment correlation tests revealed that age and Stroop score (r=–0.28, p<.001), and age and WM capacity (r=–0.45, p<.001) correlated negatively. Stroop and WM scores correlated positively (r=0.34, p<.001).

---

[2] The order of the levels reflects our hypothesis that relative to baseline (no cues), mismatching cues would interfere with comprehension, while matching cues would facilitate comprehension.

A) Mean Accuracy by Condition by Age

B) Mean Response Times by Condition by Age

Visual Cue   Congruent   Baseline   Incongruent
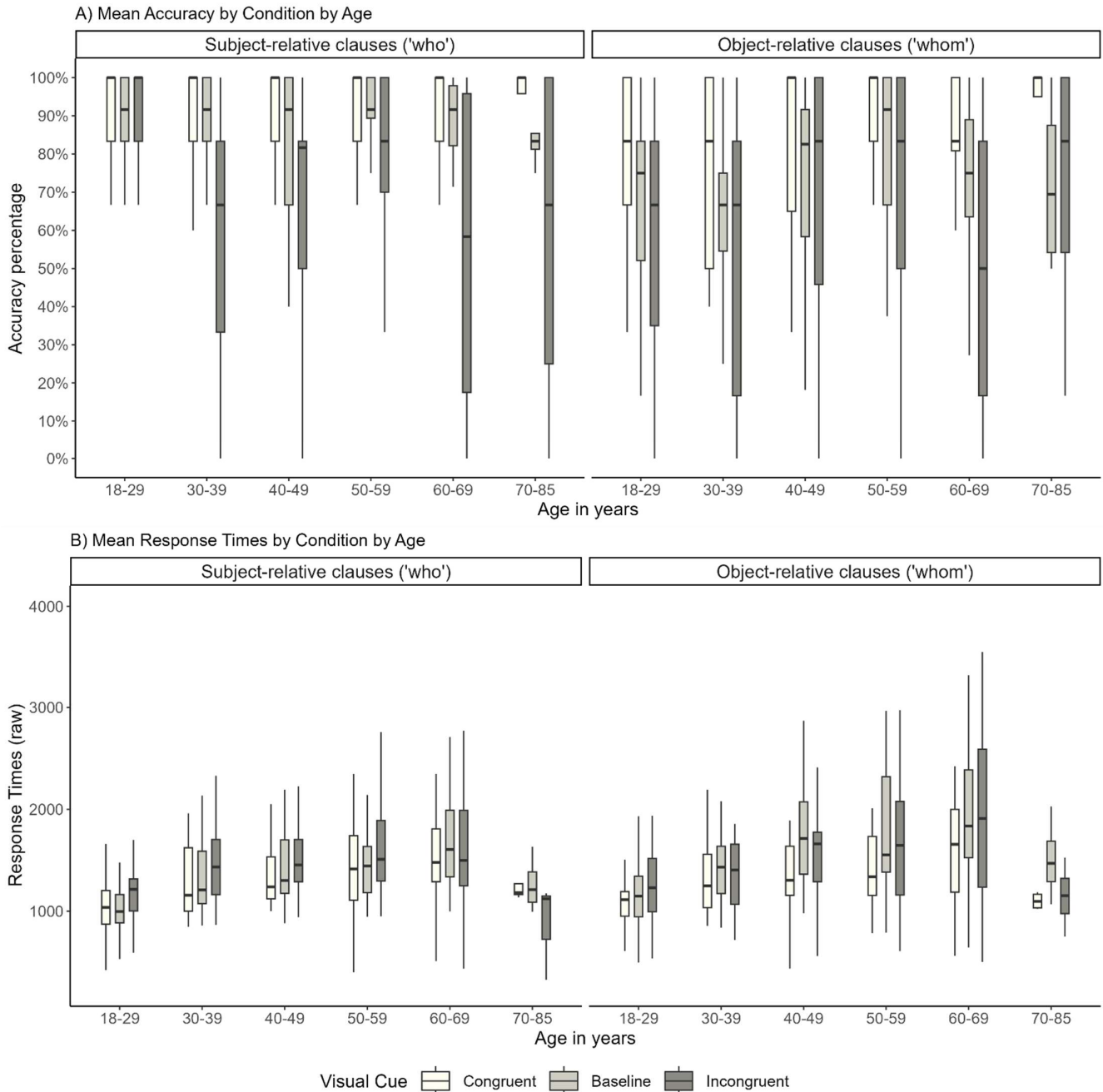
Fig 2 Mean accuracy and RTs (in milliseconds) by condition and age (on a by-participant basis). Boxes correspond to the interquartile range (IQR; 25-75%). Whiskers correspond to minimal and maximal datapoints (excluding outliers). Lines inside the boxes correspond to the condition median

5.5.2   Inferential statistics

5.5.2.1   Accuracy

The random-effects structure included random by-participant intercept and Sentence-Structure slope adjustments, and random by-item and by-task-order intercept adjustments. Task order had a significant effect as a fixed factor, as accuracy was overall higher when the

critical block was presented before (vs. after) the baseline block (the model's output is available in Table 2bis, Appendix I). To filter out the latter effect statistically without overfitting the model, task order was added in the random-effects structure. The effects of the remaining fixed factors were unchanged. Logistic regression coefficients are reported in Table 2 and plotted in Figure 3.

Table 2: Logistic regression coefficients: Accuracy data. Sentence Structure is coded as −0.5 for SR and 0.5 for OR. Visual Cue (Match vs. Baseline) is coded as −0.66 for Match and as +0.33 for Baseline and Mismatch. Visual Cue (Mismatch vs. Baseline) is coded as −0 .33 for Match and Baseline and as +0.66 for Mismatch.

Predictor	Est.	SE	z-value	p-value
(Intercept)	1.56	0.25	6.16	<.001
Visual Cue (Match vs. Baseline)	**-0.74**	**0.10**	**-7.44**	**<.001**
Visual Cue (Mismatch vs. Baseline)	**-1.01**	**0.09**	**-11.70**	**<.001**
Sentence Structure	**-0.84**	**0.12**	**-7.31**	**<.001**
Age	0.01	<0.01	-1.22	.224
Visual Cue (Match vs. Baseline) × Sentence Structure	-0.10	0.18	-0.57	.571
Visual Cue (Mismatch vs. Baseline) × Sentence Structure	**0.59**	**0.15**	**4.04**	**<.001**
Sentence Structure × Age	**0.02**	**0.01**	**3.22**	**.001**
Visual Cue (Match vs. Baseline) × Age	-0.01	0.01	-1.33	.182
Visual Cue (Mismatch vs. Baseline) × Age	**-0.02**	**<0.01**	**-3.95**	**<.001**
Visual Cue (Match vs. Baseline) × Sentence Structure × Age	-0.01	0.01	-1.16	.247
Visual Cue (Mismatch vs. Baseline) × Sentence Structure × Age	0.01	0.01	0.79	.432

Random effects: Accuracy model				
Groups	**Name**	**Variance**	**SD**	**Corr**
Subject	(Intercept)	0.80	0.89	
	Sentence Structure	1.01	1.01	0.05
Item	(Intercept)	0.06	0.25	
Task order	(Intercept)	0.11	0.33	

Formula: Accuracy ~ Sentence Structure * Visual Cue * Age + (1 + Sentence Structure | Subject) + (1 | Item) + (1 | Task Order)

The Visual Cue manipulation affected accuracy rates: relative to baseline, matching visual cues were associated with significantly higher accuracy, while mismatching visual cues were associated with lower accuracy. Subject-relative (SR) sentences yielded lower accuracy than object-relative (OR), as evidenced by a main effect of Sentence Structure. An interaction of Visual Cue (Mismatch vs. Baseline) × Sentence Structure (Figure 3) indicated that mismatching trials yielded low accuracy rates, with smaller benefits from canonical structures as compared to Baseline trials, in which the effects of Sentence Structure were statistically more pronounced. Follow-up tests showed that the Mismatch effect in response to SR (b=1.32, SE=0.12, z=10.94, p<.001) was almost double in size as compared to OR trials (b=0.72, SE=0.11, z=6.78, p<.001).
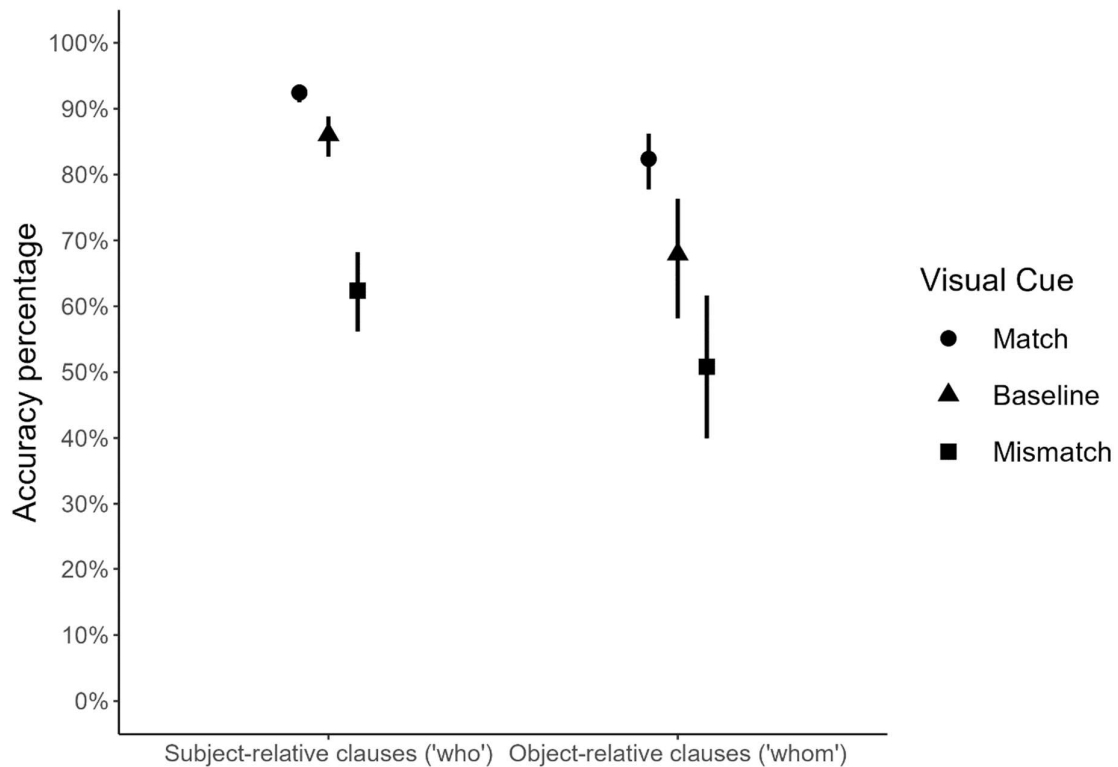
Fig 3 Sentence comprehension accuracy (y-axis) as a function of Sentence Structure (x-axis) and Visual Cue (circle: Match, triangle: Baseline, square: Mismatch). All figures display partial effects, that is, the effects of interest while holding all other continuous predictors constant at their means, and as the average of the effects of interest at different levels of other categorical variables. Error bars represent 95% confidence intervals.

There was no significant main effect of Age on overall accuracy rates. However, Age significantly interacted with Sentence Structure: the effect of Sentence Structure (greater accuracy for SR vs. OR clauses) significantly *decreased* with increasing age (Figure 4a). Follow-up analyses showed that with increasing age, accuracy decreased in SR trials (b=1.06, SE=0.39, z=2.73, p=.006), but not in OR trials (b=-0.27, SE=0.37, z=-0.77, p=.442).

Further, the interaction of Visual Cue (Mismatch vs. Baseline) × Age reflected that increasing age did not affect accuracy rates in Baseline trials (b=0.17, SE=0.34, z=0.52, p=.605), while it was associated with accuracy decreases in Mismatch trials (b=1.26, SE=0.35, z=3.56, p<.001); see Figure 4b.

A)

B)
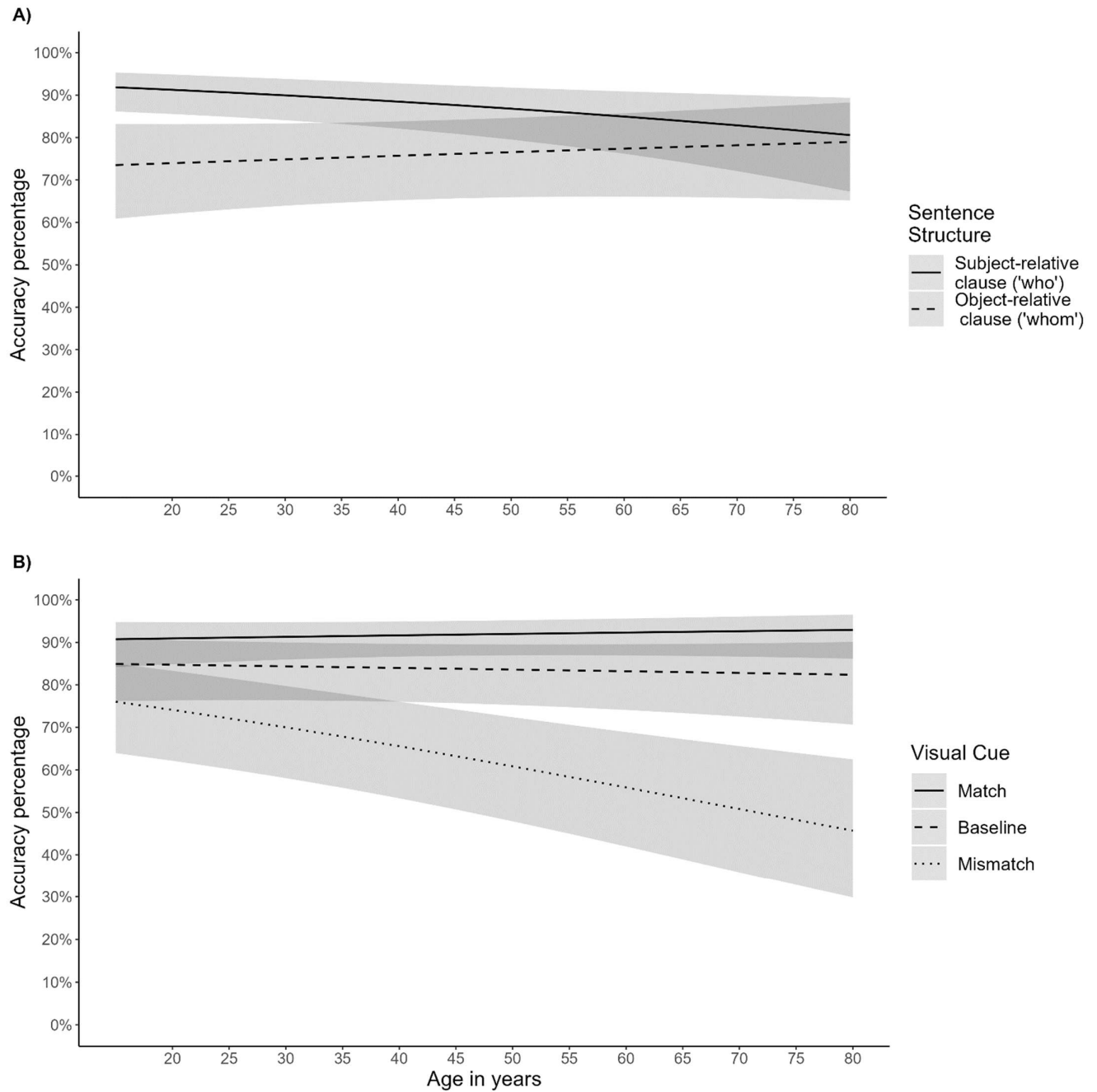
Fig 4 Shaded bands represent pointwise standard errors (95% confidence intervals are approximately twice the width of standard error bands). Panel A) Interaction of Age and Sentence Structure on sentence comprehension accuracy (solid lines: SR clauses; dashed lines: OR clauses). Panel B) Performance at sentence comprehension accuracy as a function of Age and Visual Cue (solid lines: Match, dotted lines: Baseline, dashed lines: Mismatch)

5.5.2.2   Reaction times

The output of the model fitted to the RT data is reported in Table 3. The random-effects structure included random intercepts and Sentence-Structure and Visual-Cue random slopes for participants, as well as random intercepts for items. Task order had no effect and was not considered in this analysis (see Table 3bis, Appendix I for the output of the model including Task Order). *P*-values were obtained from *t*-tests (degrees of freedom = difference between the number of observations and the number of fixed effects estimates in the model, Baayen et al., 2008).

Table 3: Linear regression coefficients: RT data. This analysis only included RTs to correctly answered trials.

Predictor	Est.	SE	t-value	p-value
(Intercept)	7.16	0.03	265.90	<.001
Visual Cue (Match vs. Baseline)	**0.14**	**0.02**	**6.77**	**<.001**
Visual Cue (Mismatch vs. Baseline)	0.05	0.03	1.91	.056
Sentence Structure	**0.07**	**0.01**	**6.12**	**<.001**
Age	**0.01**	**<0.01**	**4.78**	**<.001**
Visual Cue (Match vs. Baseline) × Sentence Structure	**0.10**	**0.02**	**4.86**	**<.001**
Visual Cue (Mismatch vs. Baseline) × Sentence Structure	**-0.07**	**0.02**	**-2.98**	**.003**
Visual Cue (Match vs. Baseline) × Age	**<0.01**	**<0.01**	**3.00**	**.003**
Visual Cue (Mismatch vs. Baseline) × Age	**<0.01**	**<0.01**	**-2.61**	**.009**
Sentence Structure × Age	<0.01	<0.01	0.93	.351
Visual Cue (Match vs. Baseline) × Sentence Structure × Age	**<0.01**	**<0.01**	**2.04**	**.041**
Visual Cue (Mismatch vs. Baseline) × Sentence Structure × Age	<0.01	<0.01	0.49	.623

Random effects: Reaction-time model					
Groups	**Name**	**Variance**	**SD**	**Corr**	
Subject	(Intercept)	0.09	0.30		
	Sentence Structure	0.01	0.08	.33	
	Visual Cue (Match vs. Baseline)	0.05	0.22	-.32	.07
	Visual Cue (Mismatch vs. Baseline)	0.09	0.30	.42	.14 -.95
Item	(Intercept)	<0.01	0.05		
Residual		0.10	0.32		

Formula: logRT ~ Sentence Structure * Visual Cue * Age + (Sentence Structure + Visual Cue | Subject) + (1 | Item)

The Visual Cue manipulation resulted in longer RTs in Baseline versus Match conditions, and marginally in the Mismatch versus Baseline condition. This result is highly compatible with the patterns obtained from the accuracy data, indicating that mismatching visual cues hindered comprehension, and that matching cues supported it in terms of both response accuracy and latency. As expected, subject-relative (SR) clauses yielded shorter RTs than object-relative (OR) clauses, and increasing age was associated with slower responses.

Further, two two-way interactions between Visual Cue and Sentence Structure indicated that the difference between subject- and object-relative clauses was significantly more pronounced in Baseline trials than in either matching or mismatching trials. In other words, Baseline trials were responded to more similarly to Mismatch (than Match) trials in response

to noncanonical OR (vs. SR) clauses (Figure 5). Follow-up analyses revealed that RTs were equally long in OR Baseline and Mismatch trials (b=-0.02, SE=0.03, z=-0.58, p=.832), while they differed significantly in response to SR trials (b=-0.08, SE=0.03, z=-2.99, p=.010). In turn, the difference in RTs between Match and Baseline trials doubled in size from SR (b=-0.09, SE=0.02, z=-4.09, p<.001) to OR conditions (b=-0.19, SE=0.02, z=-8.14, p<.001).
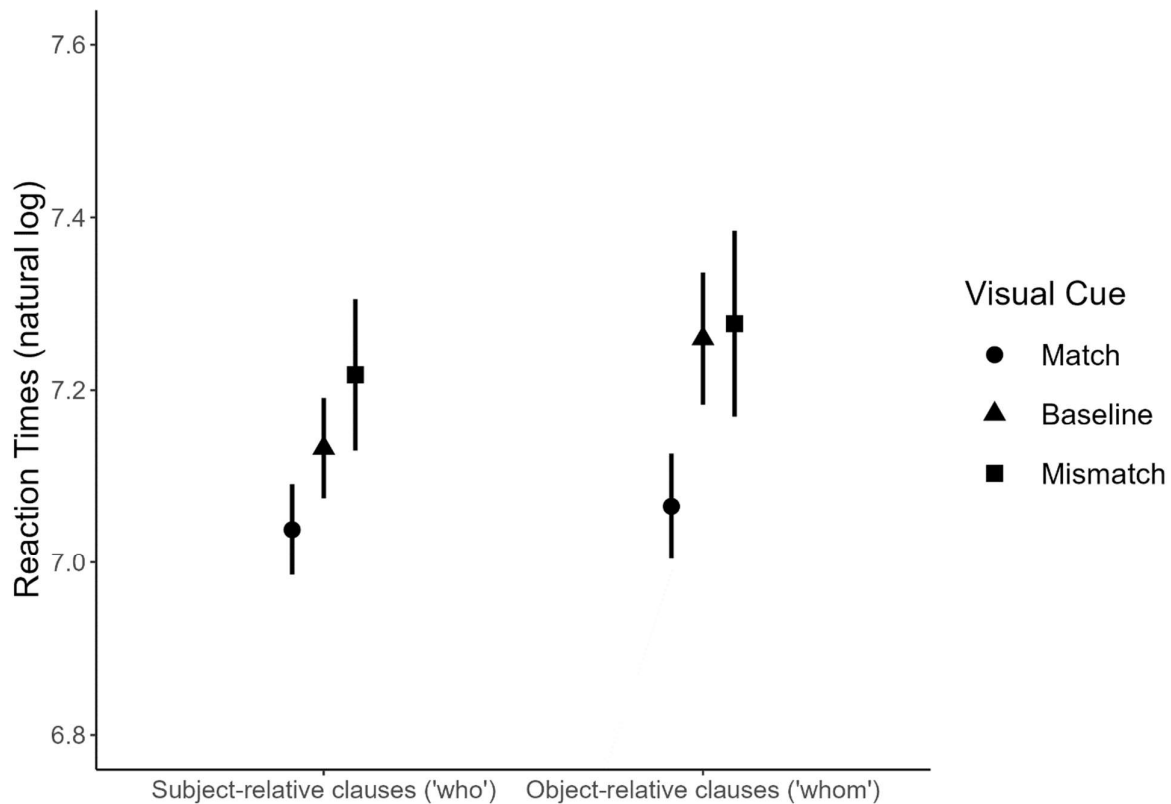

Fig 5 Sentence comprehension RTs (y-axis) as a function of Sentence Structure (x-axis) and Visual Cue (circle: Match, triangle: Baseline, square: Mismatch).

Two-way interactions between Visual Cue and Age indicated that the observed age-related slowdowns were significantly steeper in response to (no-cue) Baseline trials than for either matching or mismatching trials. To investigate this, we first limited the analysis to trials presented with matching versus mismatching visual cues (sum-coded; matching: –0.5; mismatching: +0.5). We found that Visual Cue did not interact with Age (b<0.01, SE<0.01, t=-0.48, p=.632; see Table 4, Appendix I). Then, we re-coded the factor Visual Cue as Baseline (+0.5) versus Visual-Cue trials (i.e., a subset including Match and Mismatch conditions, both coded as –0.5). The effect of Age yielded a significant interaction with Visual Cue (b<0.01 SE<0.01, t=-3.61, p<.001, Figure 6). This indicated that increasing age was associated with significantly larger increases in RTs when no visual cues were available, as compared to trials in which such cues were presented, regardless of whether they matched the role-relations

conveyed by the sentence. We discuss potential explanations of this finding in the following section.

Finally, a three-way interaction indicated that SR-OR differences (OR slower than SR) decreased with increasing age in response to Match trials, while they increased with increasing age in response to Baseline trials. This result may be interpreted as evidence for the supporting role of congruent visual scenes for comprehension in older adults. The same three-way interaction was significant in the model comparing Match and Mismatch conditions directly (see Table 4, Appendix I), suggesting that Baseline and Mismatch conditions showed comparable patterns, which differed significantly from Match conditions.
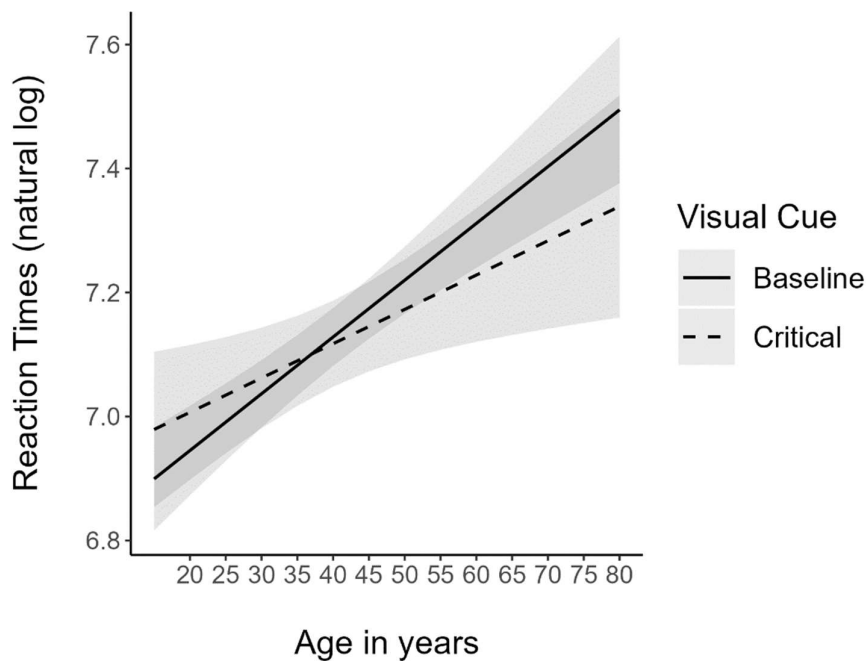
Fig 6 Sentence comprehension RTs as a function of whether Visual Cues were present (dashed line) or absent (solid line) over the age range

5.6   Discussion

The present study examined how language users across the adult lifespan rely on visual and linguistic cues for the comprehension of grammatically unambiguous sentences in German. We asked the following research questions:

1.   Do visual cues influence reliance on word-order versus morphosyntactic cues for comprehension of grammatically unambiguous sentences?

2.   Are there age-related differences in the degree to which visual cues influence reliance on word-order versus morphosyntactic cues for comprehension?

The novel results of the study can be summarized as follows: the presence of matching (vs. no) visual scenes facilitated sentence comprehension, and the presence of mismatching (vs. no) visual scenes hindered sentence comprehension, as measured by both accuracy and RTs. Noncanonical (vs. canonical) sentences yielded slower responses in the absence (vs. presence) of visual scenes. Increasing age was associated with (i) accuracy decreases in response to canonical sentences only, (ii) decreasing accuracy in trials in which the spoken sentence and the visual scene mismatched, and (iii) steeper RT increases in the absence (vs. presence) of visual scenes.

5.6.1   Interpretation of results

Regarding the first research question, we detected significant effects of both matching and mismatching scene depictions (vs. no scene depictions in the Baseline condition) on accuracy rates, regardless of Sentence Structure and across the age range. The observed patterns corresponded to our hypothesis that scene depictions that matched the thematic roles conveyed by the sentence may be beneficial, while mismatches may be detrimental for sentence comprehension. In line with predictions from the CIA, our results indicated that the abstract role-relations in the visual scenes were encoded, verified against, and integrated with language-driven representations of role-relations, thus influencing the recall of the thematic roles of the sentence. These results add to previous work by providing evidence for the integration of visual and spoken input for thematic-role assignment of unambiguous syntax (e.g., Knoeferle & Crocker, 2006), and by measuring visual effects with a linguistic task unrelated to the visual component (e.g., Knoeferle & Crocker, 2005). Further, we found that matching (vs. no) scenes neutralized both RT increases and accuracy decreases in response to noncanonical (vs. canonical) sentences; in other words, the presence of a scene depiction that was consistent with the sentence canceled out the detrimental effect of a noncanonical word order. This is in line with previous evidence indicating that converging visual cues may neutralize the agent-before-preference and dampen the difficulty associated with the processing of noncanonical syntax (e.g., Knoeferle & Kreysa, 2012; Kreysa et al., 2018). Visual scenes conveying thematic relations congruent with an OR clause may have increased the perceived likelihood of a specific referent taking on the patient role, though it was mentioned as first in the utterance.

Turning to the lifespan perspective, we observed overall age-related declines in response speed, while age had no effect on overall agent-identification accuracy. In contrast to the hypothesis that increasing age might increase vulnerability to conflicting case and word-order cues, our results showed that older (vs. younger) adults were as accurate as their younger counterparts in response to OR clauses. As laid out in the Introduction, findings for age declines in sentence comprehension tasks have been mixed. While some studies have reported age-related vulnerability in the processing of noncanonical sentences (Peelle et al., 2010; Stine-

Morrow et al., 2000), others have not (Amichetti et al., 2016; Cutter et al., 2021). The lack of age-related accuracy declines in OR trials in our study converges with the findings from Davis et al. (2014) and Samu et al. (2017), who showed that older and younger adults were equally able to revisit previously preferred interpretations upon encountering disambiguating agreement cues. In their design, participants performed an acceptability-judgment task of initially ambiguous sentences (e.g., *Landing planes… is very easy/are very noisy*). Similarly, our participants heard sentences which presented an initial bias towards the canonical subject-object reading (up to the relative pronoun), which was either supported or disconfirmed by the case cues on the relative pronoun. Like Samu et al. (2017) and Davis et al. (2014), we found no age differences in participants' ability to overturn a previously held interpretation preference upon encountering disambiguating syntactic cues.

However, we did detect age-related comprehension declines in response to canonical SR clauses. Post-hoc, a possible explanation for this counterintuitive finding may be that older (vs. younger) adults have been found to engage in predictive processing to a smaller extent (e.g., Federmeier, 2007; Federmeier et al., 2010). In our study, if older (vs. younger) adults held less strong beliefs as to the continuation of the sentence, encountering a noncanonical or a canonical structure should have come as equally (un)expected. In contrast, encountering a canonical (vs. noncanonical) structure may have been more beneficial to younger adults, who might rely on predictions from word-order cues to a greater extent than older adults.

As regards Research Question 2, increasing age was associated with greater interference from mismatching (vs. no) visual cues overall. This finding is broadly compatible with dual-route models of sentence comprehension, including the Good-Enough approach (Ferreira, 2003), and more specifically with prior work showing an increased reliance on meaning-based heuristics for sentence comprehension with increasing age (Christianson et al., 2006; Federmeier et al., 2010; Yoon, 2015). Importantly, the current evidence adds to previous work by providing evidence for the role of visual information as a non-syntactic constraint which may be increasingly relied on for comprehension at older age, alongside semantics (e.g., Beese et al., 2019). It has been proposed that such "syntax-to-semantics" shift might be caused by age-related changes in syntactic segmentation (i.e., the ability to mentally manipulate sentence constituents, e.g., by chunking and binding elements to form representations for later recall; Stine-Morrow & Payne, 2016), which might engender increased reliance on semantic and contextual information to compensate for weaker syntactic representations in delayed comprehension measures. Under this perspective, our findings might speak for a role of interfering visually driven representations during the binding of thematic-role relations in short-term memory.

Overall, these findings are in line with our hypothesis for a greater reliance on visual information in older than younger adults. However, we did not find evidence in favor of the

idea that such an age-related shift in heuristic strategies is necessarily due to difficulty in parsing noncanonical clause structures.

Finally, we turn to the finding that age-related RT increases were steeper when visual cues were absent (vs. when they were present), regardless of whether these cues matched or mismatched with the linguistically encoded action direction. This contrasted with our prediction that increasing age might be associated with disproportionate comprehension difficulty in response to mismatching conditions specifically. One possible explanation of this finding relates to facilitation of lexical processing. By presenting depictions that always included both mentioned referents performing the same action as in the sentence, matching and mismatching visual cues alike supported lexical retrieval, as the referents could be quickly accessed visually.[3] Thus, the finding that (in older adults) visual cues yielded faster responses, regardless of whether they matched the thematic roles, might be due to facilitation specific to lexical retrieval processes (see also Altmann & Kemper, 2006; Davidson et al., 2003). This explanation seems plausible due to the specificity of the visual-cue facilitation to older adults: because lexical retrieval is less problematic for younger than older adults (e.g., Schmitter-Edgecombe, 2000; Reifegerste et al., 2022), we may assume that support from visual depictions for lexical retrieval would be reflected particularly on older adults' response speed. Some corroboration for the latter explanation comes from evidence on syntactic planning and language production in healthy aging (Hardy et al., 2020), which suggests that difficulties attributed to syntactic parsing may in fact be partially rooted in age-related lexical retrieval difficulties. Alternatively, a possible explanation of this result may be based on the conclusions from a study examining age differences in the choice of strategies to solve arithmetic problems (Anders et al., 2018). The authors reported that the choice of an approximation-based (vs. computation-based) strategy was associated with decreased RTs. If we consider the older adults' increased rates of incorrect responses in Mismatch trials in the current experiment, we may speculate that choosing a heuristic route (which may have been preferred by older participants in Mismatch conditions especially) might have (relatively) increased the speed at which participants provided their response by lowering the 'response threshold' (i.e., the level of certainty needed by the participant before providing a response, Anders et al., 2018).

In conclusion, our results align with previous evidence for the role of visual information on the comprehension of unambiguous syntax and speak to the robustness of an underlying cross-modal verification mechanism, in line with theoretical accounts such as the CIA. Further, we found an increased reliance on non-structural information during sentence

---

[3] See Miklashevsky et al. (under review), Reifegerste, Meyer et al. (2021), and Pulvermüller et al. (2001, 2005, 2009) for evidence in favour of automatic lexico-semantic access even during tasks which do not require semantic processing (e.g., lexical decision), including in aging.

comprehension in healthy aging. The observed age changes in the weighting of the visual component did not seem to be attributable to an increased difficulty in overturning previously held interpretation preferences in older adults, as the age-related increases in the size of the visual effects did not seem to be affected by sentence structure (canonical, noncanonical).

5.6.2   Limitations and future directions

The present study has some limitations, which suggest additional future directions for research. First, the participants in our sample were relatively young, and further investigations should pursue the possibility that more dramatic age-related changes in processing capacities for syntactic parsing might result in comprehension declines at older ages.

Further, as the present experiment was conducted remotely during the COVID-19 pandemic, a lab-based replication should be conducted to investigate whether the present results are in part due to testing effects, especially relating to older participants' comfort in using technology, as well as to potential differences in older and younger adults' technological equipment at home. Due to the remote administration, we also had to rely on participants' self-reported visual and hearing acuity. While performance in the Baseline condition (during which participants could rely only on the auditory information) would likely have identified participants with major hearing loss, hearing acuity has been shown to play an important role in determining the relative weighting of specific cues during sentence comprehension, likely due to perceived cue reliability (e.g., Amichetti et al., 2016; Tun et al., 2010; Wingfield et al., 2015). Future research can elucidate which perceptual difficulties might contribute to shaping individual preferences for specific heuristic strategies during language comprehension.

Lastly, future research should seek to establish the extent to which difficulties with lexical processing might underlie a) age-related changes in syntactic parsing of highly complex language and b) age-related shifts in heuristic strategies for comprehension. Yet another step in this direction would be to determine whether increased reliance on visual depictions for comprehension might be causally (and/or strategically) related to older adults' (potential) use of visual-cue information for lexical processing. Speculatively, older participants might have relied on visual cues more than younger participants in our study because, contrary to our assumptions, relying on visual cue information represented an efficient strategy for better performance in the face of specific age-related lexical-retrieval deficits. Such an explanation would support the view that age-related shifts in heuristic strategies for comprehension might be self-regulated and based on considerations of effort and attention allocation, as suggested by Stine-Morrow (2007).

5.7 Appendix I

Table 2bis: Logistic regression coefficients: Accuracy data. Sentence Structure is coded as −0.5 for SR and 0.5 for OR. Task presentation order is coded as −0.5 for Baseline > Visual Cue and 0.5 for Visual Cue > Baseline.

Predictor	Est.	SE	z-value	p-value
(Intercept)	1.55	0.10	15.67	<.001
Visual Cue (Match vs. Baseline)	**-0.73**	**0.10**	**-7.40**	**<.001**
Visual Cue (Mismatch vs. Baseline)	**-1.01**	**0.09**	**-11.67**	**<.001**
Sentence Structure	**-0.84**	**0.12**	**-7.27**	**<.001**
Age	0.01	<0.01	-1.34	.178
Task presentation order (Baseline > Visual Cue)	**-0.71**	**0.17**	**-4.17**	**<.001**
Visual Cue (Match vs. Baseline) × Sentence Structure	-0.10	0.18	-0.57	.571
Visual Cue (Mismatch vs. Baseline) × Sentence Structure	**0.59**	**0.15**	**4.03**	**<.001**
Sentence Structure × Age	**0.02**	**0.01**	**3.20**	**.001**
Visual Cue (Match vs. Baseline) × Age	-0.01	0.01	-1.33	.183
Visual Cue (Mismatch vs. Baseline) × Age	**-0.02**	**<0.01**	**-3.94**	**<.001**
Visual Cue (Match vs. Baseline) × Sentence Structure × Age	-0.01	0.01	-1.15	.248
Visual Cue (Mismatch vs. Baseline) × Sentence Structure × Age	0.01	0.01	0.78	.432

Random effects: Accuracy model				
Groups	**Name**	**Variance**	**SD**	**Corr**
Subject	(Intercept)	0.78	0.89	
	Sentence Structure	1.02	1.01	0.06
Item	(Intercept)	0.06	0.25	

Formula: Accuracy ~ Sentence Structure * Visual Cue * Age + Task Order + (1 + Sentence Structure | Subject) + (1 | Item)

Table 3bis: Linear regression coefficients: RT data. This analysis only included RTs to correctly answered trials.

Predictor	Est.	SE	t-value	p-value
(Intercept)	7.00	0.04	152.63	<.001
Visual Cue (Match vs. Baseline)	**0.07**	**0.04**	**1.95**	**<.001**
Visual Cue (Mismatch vs. Baseline)	0.16	0.05	3.26	.056
Sentence Structure	**0.07**	**0.02**	**3.13**	**<.001**
Age	**0.01**	**<0.01**	**4.43**	**<.001**
Task Order	-0.06	0.04	-1.59	.855
Visual Cue (Match vs. Baseline) × Sentence Structure	**0.10**	**0.02**	**4.96**	**<.001**
Visual Cue (Mismatch vs. Baseline) × Sentence Structure	**-0.05**	**0.02**	**-2.23**	**.003**
Visual Cue (Match vs. Baseline) × Age	**<0.01**	**<0.01**	**2.56**	**.003**
Visual Cue (Mismatch vs. Baseline) × Age	**<0.01**	**<0.01**	**-2.45**	**.009**
Sentence Structure × Age	<0.01	<0.01	0.51	.351
Visual Cue (Match vs. Baseline) × Sentence Structure × Age	**-0.06**	**0.03**	**-1.97**	**.041**
Visual Cue (Mismatch vs. Baseline) × Sentence Structure × Age	0.04	0.04	1.01	.624

Random effects: Reaction-time model					
Groups	**Name**	**Variance**	**SD**	**Corr**	
Subject	(Intercept)	0.09	0.30		
	Sentence Structure	0.01	0.08	.33	
	Visual Cue (Match vs. Baseline)	0.05	0.22	-.32	.07
	Visual Cue (Mismatch vs. Baseline)	0.09	0.30	.42	.14	-.95
Item	(Intercept)	<0.01	0.05		
Residual		0.10	0.32		

Formula: logRT ~ Sentence Structure * Visual Cue * Age + Task Order + (Sentence Structure + Visual Cue | Subject) + (1 | Item)

Table 4: Linear regression coefficients: RT data. This analysis only included RTs to correctly answered trials and compared Match (–0.5) to Mismatch (+0.5) conditions directly.

Predictor	Est.	SE	t-value	p-value
(Intercept)	7.00	0.04	243.36	<.001
Visual Cue (Match vs. Mismatch)	**0.12**	**0.02**	**5.09**	**<.001**
Sentence Structure	**0.04**	**0.01**	**2.70**	**.007**
Age	**0.01**	**<0.01**	**3.33**	**.001**
Visual Cue (Match vs. Mismatch) × Sentence Structure	0.03	0.02	1.15	.248
Visual Cue (Match vs. Mismatch) × Age	<0.01	<0.01	-0.48	.632
Sentence Structure × Age	<0.01	<0.01	0.43	.670
Visual Cue (Match vs. Mismatch) × Sentence Structure × Age	**<0.01**	**<0.01**	**2.26**	**.024**

Random effects: Reaction-time model					
Groups	**Name**	**Variance**	**SD**	**Corr**	
Subject	(Intercept)	0.12	0.34		
	Sentence Structure	0.01	0.11	.16	
	Visual Cue (Match vs. Mismatch)	0.01	0.11	-.57	.41
Item	(Intercept)	<0.01	0.04		
Residual		0.10	0.31		

Formula: logRT ~ Visual Cue * Sentence Structure * Age + (Sentence Structure + Visual Cue | Subject) + (1 | Item)

6   General Discussion

This dissertation examined the contribution of (incongruent) cues from visual scenes, word order, and case, to building thematic-role representations on-line and post-sentence. The following questions were addressed:

1. How do visual scenes influence thematic-role representations of canonically and noncanonically ordered clauses on-line (Publication 1) and post-sentence (Publications 2, 3)?

2. How do visual scenes influence thematic-role representations in L1 and L2 comprehension? (Publication 2)

3. How do visual scenes influence thematic-role representations across the adult lifespan? (Publication 3)

The main findings can be summarized as follows:

1. Regarding question 1, the incongruence (vs. congruence) of visual scenes and word order yielded later signatures of thematic-role prediction, and the incongruence of visual scenes and case yielded slower and less accurate agent identification. This supports the role of thematic incongruence of visual and spoken components of the input on thematic-role representations both on-line and post-sentence.

2. As per question 2, L2 (vs. L1) speakers exhibited less reliance on case cues at the group level for comprehension of sentences in isolation. In turn, individual reliance on case in isolation (rather than L1-L2 grouping) predicted reliance on visual scenes. This corroborates the view that there are graded differences in the reliance on cues from distinct sources, and that there may be a trade-off in the reliance on case versus visual scenes for comprehension at the individual level.

3. Turning to question 3, increasing age was associated with higher reliance on visual scenes, and equal reliance on case cues. This suggests that age-related changes in comprehension may be due to enhanced conflict from incongruent visual scenes, rather than from word order, in line with the view that older (vs. younger) adults may exploit inference-based comprehension strategies to a greater extent.

In the following sections, I discuss each point in turn.

6.1   The recent-role preference

The current thesis hypothesized and tested a recent-role preference – i.e., a tendency to re-assign recently inspected thematic roles to the same referents in temporally coordinated utterances. This hypothesis was formulated based on previous evidence for the time course of integrating thematically incongruent representations from visual and linguistic modalities

(Knoeferle et al., 2014; Knoeferle & Crocker, 2005). Within the CIA, the concept of reference includes semantic entities (e.g., characters), their actions, and the relation between entities established by the action (Knoeferle et al., 2005). The representations encoded within temporally coordinated linguistic expressions are co-indexed and compared to the representations encoded from the visual scene. When an incongruence of visual and linguistic representations is detected during the co-indexing stage, visual representations may be marked as "false" (Knoeferle et al., 2014). I investigated the effects of such "false" representations on thematic-role representations of canonically and noncanonically ordered clauses.

The results showed reliable and immediate effects of visually conveyed thematic relations on predictive thematic-role comprehension. This was signaled by mouse trajectories heading towards the recently inspected agent as early as ~300ms after trial onset. This suggests that, in the face of yet uninhibited linguistic contexts (such as "This is…"), the recently inspected agent is perceived to be the most likely referent to take on the agent role in the temporally coordinated utterance. As considered by Knoeferle et al. (2005), it may be argued that this early effect reflects a visual-search stage, rather than a thematic-role preference. Although eye and mouse movements behave similarly, mouse movements are slower (Egner et al., 2018) and "broadly lower" in number than eye movements (Milisavljevic et al., 2021). This would suggest that the simple visual search necessary to identify the response buttons in Publication 1 may constrain mouse movements at trial onset only to a limited extent.

When the inspected agent was the first-mentioned referent in the utterance, the thematic-role representations conveyed by visual scenes and subject advantage converged. This resulted in anticipatory mouse movements towards the recent agent, which were initiated ~700ms before the disambiguating information was presented. This indicates that the perceived likelihood of the recent agent taking on the agent role was further enhanced by congruent word-order cues. In contrast, the (temporary) incongruence of visual-scene and word-order cues delayed the thematic-role predictions – i.e., the participants were less prone to assume that the first mentioned referent was the agent of the utterance if the same referent was the most recent patient.

Turning to the effects of sentence-picture congruence post-sentence, congruent trials yielded faster and more accurate responses than spoken sentences in isolation, and more accurate responses than incongruent trials (the response times were not statistically examined). The convergence of information from across perceptual modalities may have enabled participants to build more richly detailed mental representations of the thematic relations. From a constraint-based perspective, the likelihood of thematic relations endorsed by both visual and case cues may be perceived as higher both relative to the likelihood of case-driven relations incongruent with the scene, and to the likelihood of relations supported by case alone. Further, the effects of congruent scenes were observed regardless of clause

structure. This implies that visually conveyed thematic-role representations were able to override syntactic preferences, such as the subject advantage, again suggesting that comprehenders were prone to assume that the recently inspected agent would be the agent of the temporally coordinated utterance, regardless of whether it was mentioned first.

Incongruent scenes yielded less accurate responses both as compared to congruent scenes (Publication 2) and to sentences in isolation (Publication 3). In Publication 1, response times were longer in response to incongruent (vs. congruent) scenes, but the accuracy data only reflected a numerical effect of sentence-picture incongruence. Here, I advance three (mutually compatible) explanations for this discrepancy.

Firstly, Publication 1 presented the visual scenes in Step 1 of the trial, and the task was always explicitly to click on the agent as the spoken stimulus unfolded in Step 2. In contrast, the probe questions in Publications 2 and 3 assessed agent identification implicitly. Further, the question itself (*Is the bride being followed?*) had to be parsed and introduced an additional thematic-role configuration which needed to be confirmed or rejected. This difference in the tasks may have caused accuracy rates in Publications 2 and 3 to be more diagnostic of the participants' ability to retrieve and manipulate thematic-role representations from memory as compared to Publication 1.

Secondly, the response could be provided as the utterance unfolded in Publication 1, while the written probe questions in Publications 2 and 3 were presented after the sentence had been spoken. The possibility to respond during the unfolding of the sentence may have facilitated agent identification based on the current linguistic information, rather than requiring participants to rely on stored representations. Thus, in addition to the issues relating to probe difficulty in the previous point, the time-gap differences between sentence presentation and probe across Publications may have engaged the participants' ability to retrieve and manipulate representations from memory to systematically different extents.

Both explanations so far are in line with the arguments brought forward by Bader & Meng (2018; 2023), and Meng & Bader (2021). Rather than attributing misinterpretations of thematic roles to failures in the mapping of form to meaning in the face of noncanonical, yet unambiguous syntactic structures, the authors maintain that such misinterpretations can be explained in terms of memory-retrieval failures specifically triggered by post-sentence agent-identification tasks.

Thirdly, visual and spoken stimuli were presented successively in Publication 1 and simultaneously in Publications 2 and 3. As the visual input was unavailable during the unfolding of the utterance in Publication 1, it may have no longer been sufficiently salient to influence the perceived likelihood of linguistically conveyed relations. Recall, however, that the visual scene was always inspected one step before the agent-identification probe was presented, such that the observed discrepancies across Publications cannot be attributed to

differences in the recency of the inspection. In contrast, the simultaneous presentation of both components may have further encouraged participants to co-index visual and spoken input, possibly exacerbating the effects of reconciling (co-present) incongruent representations, as compared to successive presentation. This consideration is consistent with previous evidence showing possible differences due simultaneous versus successive encoding in sentence-picture verification outcomes (e.g., Underwood et al., 2004).

To address these open questions, future studies are planned to compare the effects (i) of simultaneous versus successive encoding of spoken and visual components of the input, (ii) of their respective recency relative to the probe, as well as (iii) of varying levels of task difficulty, on the observed comprehension patterns (see below).

Taken together, the current findings provide support for the hypothesized recent-role preference – i.e., the tendency to re-assign recently inspected thematic roles to the same referents in temporally coordinated utterances – and point to its interaction with word-order cues in determining the time course of thematic-role processing. In line with the constraint-based approach, this dissertation assumed that all available cues contribute (to various extent) to determining the perceived likelihood of specific thematic-role configurations between constituents. Thus, the observed evidence for the recent-role preference is interpreted in terms of visually driven changes in the perceived likelihood for specific referents to take on specific roles currently and in the near future, irrespective of the source of the representation. This interpretation is in line both with current and past models of visuo-linguistic integration (CCM, CACTRAS, CIA), that propose that scenes and sentences are stored in a common representational format (see also Yildirim & Jacobs, 2013, 2015).

Evidence supporting the recent-role preference has previously been reported from neurophysiological (Knoeferle et al., 2014), eye-tracking (Knoeferle et al., 2005), and reading-time measures (Knoeferle & Crocker, 2005). This dissertations' findings add to Knoeferle et al.'s (2014) evidence for an incremental verification process by showing on-line and post-sentence effects of competition between visually and linguistically conveyed thematic relations during a linguistic task unrelated to sentence-picture congruence – similar to Knoeferle & Crocker (2005). In contrast to both Knoeferle & Crocker (2005) and Knoeferle et al. (2014), this dissertation used spoken stimuli. The findings support the generalizability of a truth-indexing mechanism of incongruent representations across stimuli, tasks, and measures of interest. Furthermore, the results suggest that incongruent representations may linger to affect thematic-role representations post-sentence. Together, this evidence contributes to specifying the time course, dynamics, and outcomes of integrating thematically incongruent visual and linguistic representations within the CIA.

Lastly, this dissertation kept the visual availability of referents and actions constant across conditions. This aimed to minimize the differences in the visual grounding of distinct semantic entities to single out the contribution of the grounding of thematic relations to

comprehension and processing. As such, the recent-role preference may contribute to explaining the robustness of recent-event fixations in the face of manipulations aimed to encourage predictive processing (Abashidze et al., 2019; Abashidze & Knoeferle, 2021). Namely, recent-event fixations may reflect both (i) a preference for recently inspected over non-inspected referents – as "generically relevant" semantic entities – i.e., a "recent-referent" preference, so to say, and (ii) the recent-role preference. To elucidate this, future investigations may directly compare the relative contribution of recent-role and "recent-referent" preferences to recent-event fixations within the same design.

6.2   Cue sensitivity in L1 and L2

With regard to L2 speakers, the thesis examined (i) whether L2 speakers' reliance on inference-based over combinatorial processing strategies for comprehension extends to the use of visual cues, and (ii) how the use of visual cues relates to that of word order and case in building thematic-role representations. Furthermore, it explored whether sensitivity to case information may predict reliance on visual scenes in L1 and L2 thematic-role comprehension.

L2 speakers responded less accurately to sentences in isolation, particularly to noncanonical clauses, indicating overall lower sensitivity to case cues compared to L1 speakers. This result closely replicated previous findings for L1-L2 differences in the comprehension of canonical and noncanonical structures (Jackson, 2008; Jackson & Bobb, 2009; Mitsugi & MacWhinney, 2016), supporting the view that L2 speakers rely on case cues to a smaller extent than L1 speakers and rather tend to rely on word order (Hopp, 2015b).

Furthermore, L2 and L1 speakers relied on visual scenes to the same extent on the group level. These patterns align with predictions from the SSH and extend the evidence supporting the CIA to L2 speakers. Additionally, this result extends the evidence supporting a preferential reliance on inference-based over combinatorial processing in L2 comprehension to visual-scene information. Compatible with these findings, recent research comparing recent-event fixations and item recall in L1 and early and late L2 speakers of English found no significant differences on either variable when comparing the three groups within the same model (Abashidze et al., 2023). Together, these patterns suggest that the integration of visual-scene information for L1 and L2 sentence comprehension proceeds along fundamentally similar mechanisms and with comparable time courses.

Importantly, in both L1 and L2 speakers, individual sensitivity to case predicted the size of visual effects. Participants at either end of the sensitivity range showed comparable visual effects in response to canonical sentences. In response to noncanonical sentences, the size of the visual effects was greatly increased at the lower end of the sensitivity range, while it disappeared at the higher end. This pattern supports the hypothesized negative association in the reliance on syntactic versus non-syntactic, and specifically visual cues across L1 and L2 speakers. This suggests a trade-off in the use of case and visual information for

comprehension, in line with neurophysiological and behavioral evidence showing individual-level preferences in the reliance on distinct linguistic cues (i.e., on syntactic versus non-syntactic information) for sentence processing (Hopp, 2015a; Pakulak & Neville, 2011; Tanner et al., 2014). The thesis' results extend this evidence by showing individual preferences for reliance on non-linguistic non-syntactic cues, such as visual scenes, for comprehension.

Lastly, case sensitivity was found to be a better predictor of agent-identification accuracy than the categorical L1-L2 grouping. This result points to graded, rather than dichotomous L1-L2 differences in sentence comprehension – as captured by theoretical proposals such as the SSH and Cunnings' (2017) Interference Hypothesis. Together, the results suggest that comprehension differences may be better understood and modeled empirically in terms of individual sensitivity to specific linguistic cues.

6.3   The adult lifespan

Turning to the lifespan trajectory, I hypothesized that visual cues might be considered as one of the information sources that are relied on increasingly with increasing age during sentence comprehension. This hypothesis is based on previous evidence for a shift from algorithmic processing strategies towards inference-based reasoning with increasing age.

The results supported this hypothesis, as the detrimental effects of sentence-picture incongruence on agent identification increased steeply with increasing age as compared to sentences in isolation. Conversely, age-related comprehension declines were numerically less steep in response to sentences presented with congruent scenes (vs. in isolation). In other words, the positive effects of sentence-picture congruence increased (though only numerically) with increasing age. These patterns indicate that the endorsement of both congruent and incongruent visual representations tended to increase as a function of increasing age.

These results are compatible with previous evidence for greater reliance on perceptual cues in older than in younger adults (Dijkstra et al., 2004; Madden & Dijkstra, 2010). According to Situation Model theory (Radvansky, 1998; Zwaan & Radvansky, 1998), age-related changes in comprehension strategies can be ascribed to varying abilities to make use of information inferred from situation models – i.e., from prior knowledge of situations as typically occurring in the real world. Critically, such ability is thought to be stable or improve throughout the lifespan, while other linguistic abilities may decline with increasing age. As a consequence, inferences based on visual information may become more prominent in older than in younger adults' comprehension.

By itself, increasing age was not a significant predictor of comprehension accuracy, and especially the comprehension of noncanonical sentences was maintained throughout the age range. This pattern is inconsistent with prior work showing decreased sensitivity to syntactic

markers of noncanonical structures in older adults (Stine-Morrow et al., 2000). With regard to this, it is worth considering that the age range of the sample included in the current work was relatively young overall (i.e., most older adults were around 60 years of age). Thus, it is possible that age-related declines in sensitivity to accusative case markings may become evident only later in life. In addition, age-related decreases in comprehension accuracy have often only been shown in response to clauses with dependencies spanning across several words (e.g., Beese, Vassileiou, et al., 2019; Reifegerste & Felser, 2017), which was not the case in this thesis' stimuli.

A further consideration about the observed results relates to the thesis' assumption that syntactic preferences represent constraints on predictive sentence processing, without which the effects of "unexpected" noncanonical structures on behavioral and neurophysiological outcomes would seem hard to explain (Kuperberg & Jaeger, 2016). Predictive processing has been found to play a more limited role in older than in younger adults' sentence processing (Federmeier et al., 2002; Wlotko & Federmeier, 2012). If older adults hold weaker beliefs about the first-mentioned referent being the agent or the patient of the action, this should result in lower conflict engendered upon encountering syntactic markers signaling a noncanonical structure. Speculatively, the current evidence may reflect "two sides of the same coin" (i.e., complementary aspects of the lifespan trajectory): A stronger recent-role preference in thematic-role comprehension with increasing age, alongside (what may be interpreted as) weaker reliance on the subject advantage. A testable prediction following from this interpretation would be that increasing age should be associated with increased fixations towards recently inspected (vs. potentially upcoming) referents in recent-event designs.

Interestingly, while response times increased with age overall, such increases were steeper when sentences were presented in isolation as compared to when sentences were presented alongside any visual-scene cues. This result is explained in terms of lexical facilitation, with consideration of two empirical facts. Firstly, there is evidence that lexical access is relatively automatic also in response to tasks which would not technically require it, such as lexical decision (Pulvermüller et al., 2005, 2009; Reifegerste, 2021). Secondly, performance declines have been reported in older adults' lexical processing (e.g., Reifegerste et al., 2021; Schmitter-Edgecombe, 2000). Thus, because the facilitating role of visual scenes on response times was only observed with increasing age, and although the task did not specifically require lexical processing, this pattern might be ascribed to the facilitating role of visual scenes for lexical retrieval during comprehension. This interpretation could be tested by comparing the visual scenes to baseline condition presenting both referents in a neutral position (rather than no visual display at all), similar to Knoeferle & Crocker (2005) (see below). If such "lexical-facilitation" hypothesis were to be supported, it would point to the possibility that increased reliance on visual information with increasing age may be strategically pursued to compensate for age-related changes in specific linguistic domains (cf. Stine-Morrow, 2007). In

real-world situations, the advantages of such a strategy might outweigh the costs associated with the occasional misunderstanding.

Together, the findings point to an increasingly important role of visual-scene cues in building thematic-role representations throughout the adult lifespan and support the view that visual-scene information should be considered as one of the non-syntactic cues relied upon increasingly with increasing age for sentence comprehension.

7   Conclusions, limitations, and future directions

The thesis presents (quite) a few limitations that warrant the need for further investigations on the topic. Some of these shortcomings have been brought up in the respective Publications and in the General Discussion. Here, I discuss issues relating to the thesis in its entirety.

Importantly, the design choice of presenting only masculine referents as head nouns was dictated by the need to make the relative clauses thematically unambiguous from the first constituent. In addition, this enabled participants to discern visually between both characters with ease – for example, participants might have struggled to distinguish a cook from a waiter in the depiction, while the male-female juxtaposition was arguably very salient. However, this implied that the male referents were always the agents of subject-extracted canonical clauses, and the female referents were always the agent of object-extracted, noncanonical clauses. Thus, the effects of the subject advantage observed in the current thesis are perfectly confounded with those of the "male advantage", as males are more likely to be perceived as agents in real life (Hsu et al., 2021) and in linguistic tasks (Esaulova et al., 2017; Esaulova & Stockhausen, 2015).

As mentioned, using sentences in isolation as a baseline to assess the effects of visual scenes presented both advantages and disadvantages. On the one hand, sentences in isolation were chosen to assess sensitivity to case information in Publication 2, and the simultaneous presentation of visual-scene information did not seem to be justified by the purpose of the assessment. On the other hand, in Publication 3, comparing sentences in isolation to those presented with visual scenes has certainly resulted in a less strictly matched baseline condition relative to the conditions of interest. In contrast, Knoeferle & Crocker (2005) presented a baseline condition in which the characters were depicted in a neutral position. For example, recall the interaction of age and visual scenes which showed steeper age-related increases in response times to sentences in isolation than to both congruent and incongruent trials in Publication 3. If, as tentatively proposed, such interaction were due to lexical facilitation, presenting depictions of both referents in a neutral position as baseline might have neutralized it fully. However, the design choices in the current thesis are further justified by previous evidence showing that depictions of characters in a neutral position may be perceived as a

source of incongruence relative to an utterance describing the characters engaging in an action (Knoeferle & Crocker, 2005). Instead, sentences in isolation were meant to introduce no further dimension of incongruence. An avenue for future studies may be to directly examine how thematic-role comprehension of sentences in isolation compares to that of sentences presented with depictions referents in a neutral position, and with depictions of referents as though engaging in an action.

Finally, as laid out above, the design and procedural differences between Publications 1 and Publications 2/3 make it impossible to disentangle the role of (i) implicit versus explicit agent-identification probes, (ii) probe timing, and (iii) simultaneous versus successive encoding of spoken and visual stimuli on thematic-role representations post-sentence. To address these open questions, three follow-up experiments are in preparation.

Firstly, the effects of visual scenes as measured by implicit versus explicit agent-identification tasks will be directly compared. To do so, the trials will be administered in two successive blocks. The analysis will compare (i) the explicit-probe block versus the implicit probes requiring Yes responses, and (ii) the implicit probes requiring *Yes* responses versus implicit probes requiring *No* responses. This will contribute to defining the effects of an additional thematic-role representation that is introduced and must be rejected, as compared to an established thematic-role representation to be confirmed.

Secondly, the differences yielded by varying probe timings will be assessed systematically. The previous design will be adapted to present implicit/explicit probe blocks either while the sentence is unfolding, or on the following slide. If memory retrieval is causally involved in agent-identification accuracy decreases, rather than "false" visual representations, responding to an explicit probe while the sentence is unfolding should yield comparable accuracy rates regardless of sentence-picture congruence (as found in Publication 1) and of clause type. Implicit probes should be disproportionately difficult in the delayed condition, which would suggest that the tasks in Publications 2 and 3 may have artificially emphasized the effects of visual scenes on thematic-role comprehension.

Thirdly, I will examine the effects of simultaneous versus successive encoding of spoken and visual input, while controlling for the recency of presentation relative to the probe. In simultaneous conditions, visual and spoken stimuli are presented simultaneously (for ~3000ms), then a probe is presented immediately or after a 3000ms lag. In successive conditions, spoken and visual stimuli are presented one after the other (for 3000ms each), and the probe question is presented following the second stimulus, resulting in probes presented immediately after the sentence or after a 3000ms lag. This evidence will be directly informative as to the reasons underlying the discrepancies observed in the accuracy results in Publications 1 versus 2 and 3.

In conclusion, this thesis contributed to defining the mechanisms underlying the integration of thematically incongruent case, word-order and visual-scene cues for thematic-role processing and representation within the CIA. The findings provide evidence for a tendency to re-assign recently inspected thematic roles to the same referents in temporally coordinated utterances, which I dubbed "recent-role" preference. Such preference is attributed to the visual grounding of thematic relations, which increases the perceived likelihood that the same relations will be re-instantiated in the utterance. The examination of additional populations of interest, based on theory-driven hypotheses, highlighted the role of individual-level variables such as chronological age and sensitivity to morphosyntactic markers in determining thematic-role comprehension strategies.

References

Abashidze, D., & Knoeferle, P. (2021). Influence of actor's congruent and incongruent gaze on language processing. *Frontiers in Psychology*, *12*(October), 1–14. https://doi.org/10.3389/fpsyg.2021.701742

Abashidze, D., Carminati, M. N., & Knoeferle, P. (2019). Anticipating a future versus integrating a recent event? Evidence from eye-tracking. *Acta Psychologica*, *200*(May 2018), 102916. https://doi.org/10.1016/j.actpsy.2019.102916

Abashidze, D., Knoeferle, P., & Carminati, M. N. (2014). How robust is the recent event preference? *Proceedings of the 36th Annual Meeting of the Cognitive Science Society, CogSci 2014*, 92–97.

Abashidze, D., Schmidt, A., Trofimovich, P., & Mercier, J. (2023). Integration of visual context in early and late bilingual language processing: evidence from eye-tracking. *Frontiers in Psychology*, *14*(April). https://doi.org/10.3389/fpsyg.2023.1113688

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models. *Journal of Memory and Language*, *38*(4), 419–439. https://doi.org/10.1006/jmla.1997.2558

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition*, *111*(1), 55–71. https://doi.org/10.1016/j.cognition.2008.12.005

Altmann, G. T. M., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, *33*(4), 583–609. https://doi.org/10.1111/j.1551-6709.2009.01022.x

Altmann, G. T. M., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*(3), 191–238. https://doi.org/10.1016/0010-0277(88)90020-0

Altmann, L. J. P., & Kemper, S. (2006). Effects of age, animacy and activation order on sentence production. *Language and Cognitive Processes*, *21*(1–3), 322–354. https://doi.org/10.1080/0169096054400006

Amichetti, N. M., White, A. G., & Wingfield, A. (2016). Multiple solutions to the same problem: Utilization of plausibility and syntax in sentence comprehension by older adults with impaired hearing. *Frontiers in Psychology*, *7*(MAY), 1–13. https://doi.org/10.3389/fpsyg.2016.00789

Anders, R., Hinault, T., & Lemaire, P. (2018). Heuristics versus direct calculation, and age-related differences in multiplication: an evidence accumulation account of plausibility decisions in arithmetic. *Journal of Cognitive Psychology*, *30*(1), 18–34. https://doi.org/10.1080/20445911.2017.1397676

Arnaudova, O., Browne, W., Rivero, M. L., & Stojanovic, D. (2004). Relative Clause Attachment in Bulgarian. *Annual Workshop on Formal Approaches to Slavic Languages*.

Ashaie, S., & Obler, L. (2014). Effect of age, education, and bilingualism on confrontation naming in older illiterate and low-educated populations. *Behavioural neurology*, *2014*, 970520. https://doi.org/10.1155/2014/970520

Ayasse, N. D., Hodson, A. J., & Wingfield, A. (2021). The principle of least effort and comprehension of spoken sentences by younger and older adults. *Frontiers in Psychology*, *12*(March). https://doi.org/10.3389/fpsyg.2021.629464

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bader, M., & Häussler, J. (2010). Toward a model of grammaticality judgments. *Journal of Linguistics*, *46*(2), 273–330. https://doi.org/10.1017/S0022226709990260Abashidze, D., Carminati, M. N., & Knoeferle, P. (2019). Anticipating a future versus integrating a recent event? Evidence from eye-tracking. *Acta Psychologica*, *200*(May 2018), 102916. https://doi.org/10.1016/j.actpsy.2019.102916

Bader, M., & Häussler, J. (2010). Toward a model of grammaticality judgments. *Journal of Linguistics*, *46*(2), 273–330. https://doi.org/10.1017/S0022226709990260

Bader, M., & Meng, M. (1999). Subject-object ambiguities in German embedded clauses: An across-the-board comparison. *Journal of Psycholinguistic Research*, *28*(2), 121–143. https://doi.org/10.1023/A:1023206208142

Bader, M., & Meng, M. (2018). The misinterpretation of noncanonical sentences revisited. *Journal of Experimental Psychology: Learning Memory and Cognition*, *44*(8), 1286–1311. https://doi.org/10.1037/xlm0000519

Bader, M., & Meng, M. (2023). Processing noncanonical sentences: effects of context on online processing and (mis)interpretation. *Glossa Psycholinguistics*, *2*(1), 1–45. https://doi.org/10.5070/G6011117

Barr, D. J. (2022). *Clusterperm: Cluster-permutation tests for time-series data from factorial experiments*.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–660. https://doi.org/10.1017/S0140525X99002149

Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, *59*(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using {lme4}. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bates, E., & MacWhinney, B. (1981). Second-Language Acquisition From a Functionalist Perspective: Pragmatic, Semantic, and Perceptual Strategies. *Annals of the New York Academy of Sciences*, *379*(1), 190–214. https://doi.org/10.1111/j.1749-6632.1981.tb42009.x

Bates, E., & MacWhinney, B. (1989). Functionalism and the Competition Model. In E. Bates & B. MacWhinney (Eds.), *The crosslinguistic study of sentence processing* (pp. 3–76). Cambridge University Press.

Bates, E., McNew, S., MacWhinney, B., Devescovi, A., & Smith, S. (1982). Functional constraints on sentence processing: A cross-linguistic study. *Cognition*, *11*(3), 245–299. https://doi.org/10.1016/0010-0277(82)90017-8

Beese, C., Vassileiou, B., Friederici, A. D., & Meyer, L. (2019). Age differences in encoding-related alpha power reflect sentence comprehension difficulties. *Frontiers in Aging Neuroscience*, *10*(JUL), 1–14. https://doi.org/10.3389/fnagi.2019.00183

Beese, C., Werkle-Bergner, M., Lindenberger, U., Friederici, A. D., & Meyer, L. (2019). Adult age differences in the benefit of syntactic and semantic constraints for sentence processing. *Psychology and Aging*, *34*(1), 43–55. https://doi.org/10.1037/pag0000300

Benţea, A. (2012). Does "Case" Matter in the Acquisition of Relative Clauses in Romanian? In A. Biller, E. Chung, & A. Kimball (Eds.), *BUCLD 36 Online Proceedings Supplement* (pp. 1–12).

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

Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2011). Expectations from preceding prosody influence segmentation in online sentence processing. *Psychonomic Bulletin and Review*, *18*(6), 1189–1196. https://doi.org/10.3758/s13423-011-0167-9

Campbell, K. L., Samu, D., Davis, S. W., Geerligs, L., Mustafa, A., & Tyler, L. K. (2016). Robust resilience of the frontotemporal syntax system to aging. *Journal of Neuroscience*, *36*(19), 5214–5227. https://doi.org/10.1523/JNEUROSCI.4561-15.2016

Caplan, D., & Waters, G. (2005). The relationship between age, processing speed, working memory capacity, and language comprehension. *Memory*, *13*(3–4), 403–413. https://doi.org/10.1080/09658210344000459

Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review*, *82*(1), 45–73. https://doi.org/10.1037/h0076248

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning Memory and Cognition*, *30*(3), 687–696. https://doi.org/10.1037/0278-7393.30.3.687

Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language*, *47*(1), 30–49. https://doi.org/10.1006/jmla.2001.2832

Christianson, K., Williams, C. C., Zacks, R. T., & Ferreira, F. (2006). Younger and older adults' "good-enough" interpretations of garden-path sentences. *Discourse Processes*, *42*(2), 205–238. https://doi.org/10.1207/s15326950dp4202_6

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*(1), 3–42. https://doi.org/10.1017/S0142716406060024

Clahsen, H., & Felser, C. (2018). Some Notes on the Shallow Structure Hypothesis. *Studies in Second Language Acquisition*, *40*, 1–14. https://doi.org/10.1017/S0272263117000250

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology*, *3*(3), 472–517. https://doi.org/10.1016/0010-0285(72)90019-9

Coltheart, M. (1999). Modularity and cognition. *Trends in Cognitive Sciences*, *3*(3), 115–120. https://doi.org/10.1016/S1364-6613(99)01289-9

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256. https://doi.org/10.1037/0033-295X.108.1.204

Connor, L. T., Spiro, A., Obler, L. K., & Albert, M. L. (2004). Change in object naming ability during adulthood. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, *59*(5), 203–209. https://doi.org/10.1093/geronb/59.5.P203

Cooper, R. M. (1974). *Control of Eye Fixation By the Meaning of Spoken Language. 107*, 166–168. https://doi.org/10.1016/0010-0285(74)90005-x

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge University Press. https://rm.coe.int/1680459f97 (accessed on 16 November 2022).

Crocker, M. W., Knoeferle, P., & Mayberry, M. R. (2010). Situated sentence processing: The coordinated interplay account and a neurobehavioral model. *Brain and Language*, *112*(3), 189–201. https://doi.org/10.1016/j.bandl.2009.03.004

Cunnings, I. (2017). Parsing and Working Memory in bilingual sentence processing. *Bilingualism*, *20*(4), 659–678. https://doi.org/10.1017/S1366728916000675

Cunnings, I., & Fujita, H. (2021). Similarity-based interference and relative clauses in second language processing. *Second Language Research*, *39*(2), 1–25. https://doi.org/10.1177/02676583211063534

Cunnings, I., Fotiadou, G., & Tsimpli, I. (2017). Anaphora Reolution and Reanalysis during L2 Sentence Processing. *Studies in Second Language Acquisition*, *39*(4), 621–652. https://doi.org/10.1017/S0272263116000292

Cutter, M. G., Paterson, K. B., & Filik, R. (2021). Online representations of non-canonical sentences are more than good-enough. *Quarterly Journal of Experimental Psychology*, (May), 174702182110320. https://doi.org/10.1177/17470218211032043

Davidson, D. J., Zacks, R. T., & Ferreira, F. (2003). Age preservation of the syntactic processor in production. *Journal of Psycholinguistic Research*, *32*(5), 541–566. https://doi.org/10.1023/A:1025402517111

Davis, S. W., Zhuang, J., Wright, P., & Tyler, L. K. (2014). Age-related sensitivity to task-related modulation of language-processing networks. *Neuropsychologia*, *63*, 107–115. https://doi.org/10.1016/j.neuropsychologia.2014.08.017

DeDe, G. (2015). Effects of animacy on processing relative clauses in older and younger adults. *Quarterly Journal of Experimental Psychology*, *68*(3), 487–498. https://doi.org/10.1080/17470218.2014.956766

DeDe, G., Caplan, D., Kemtes, K. A., & Waters, G. S. (2004). The relationship between age, verbal working memory, and language comprehension. *Psychology and Aging*, *19*(4), 601–616. https://doi.org/10.1037/0882-7974.19.4.601

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–1121. https://doi.org/10.1038/nn1504

den Ouden, D. B., Dickey, M. W., Anderson, C., & Christianson, K. (2016). Neural correlates of early-closure garden-path processing: Effects of prosody and plausibility. *Quarterly Journal of Experimental Psychology*, *69*(5), 926–949. https://doi.org/10.1080/17470218.2015.1028416

Deniz, N. D. (2022). Processing syntactic and semantic information in the L2: Evidence for differential cue-weighting in the L1 and L2. *Bilingualism*. https://doi.org/10.1017/S1366728921001140

Díaz, B., Erdocia, K., Menezes, R. F. de, Mueller, J. L., Sebastián-Gallés, N., & Laka, I. (2016). Electrophysiological correlates of second-language syntactic processes are related to native and second language distance regardless of age of acquisition. *Frontiers in Psychology*, *7*(February). https://doi.org/10.3389/fpsyg.2016.00133

Dijkstra, K., Yaxley, R. H., Madden, C. J., & Zwaan, R. A. (2004). The role of age and perceptual symbols in language comprehension. *Psychology and Aging*, *19*(2), 352–356. https://doi.org/10.1037/0882-7974.19.2.352

Egner, S., Reimann, S., Hoeger, R., & Zangemeister, W. H. (2018). Attention and information acquisition: Comparison of mouse-click with eye-movement attention tracking. *Journal of Eye Movement Research*, *11*(6), 1–27. https://doi.org/10.16910/jemr.11.6.4

Eiteljoerge, S. F. V., Adam, M., Elsner, B., & Mani, N. (2019). Consistency of co-occurring actions influences young children's word learning. *Royal Society Open Science*, *6*(8), 190097. https://doi.org/10.1098/rsos.190097

Elman, J. L., Hare, M., & McRae, K. (1996). Cues, constraints, and competition in sentence processing. In M. Tomasello & D. Slobin (Eds.), *Beyond nature-nurture: Essays in honor of elizabeth bates*. Lawrence Erlbaum Associates.

Esaulova, Y., & Stockhausen, L. von. (2015). Cross-linguistic evidence for gender as a prominence feature. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01356

Esaulova, Y., Reali, C., & Von Stockhausen, L. (2017). Prominence of gender cues in the assignment of thematic roles in German. *Applied Psycholinguistics*, *38*(5), 1133–1172. https://doi.org/10.1017/S014271641700008X

Farmer, T. A., Anderson, S. E., & Spivey, M. J. (2007). Gradiency and visual context in syntactic garden-paths. *Journal of Memory and Language*, *57*(4), 570–595. https://doi.org/10.1016/j.jml.2007.04.003

Farmer, T. A., Misyak, J. B., & Christiansen, M. H. (2012). Individual differences in second language sentence processing. In M. J. Spivey, M. F. Joannisse, & K. McRae (Eds.), *The Cambridge Handbook of Psycholinguistics* (pp. 353–364). Cambridge University Press. https://doi.org/10.1111/j.1467-9922.2012.00711.x

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491–505. https://doi.org/10.1111/j.1469-8986.2007.00531.x

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

Federmeier, K. D., McLennan, D. B., Ochoa, E. de, & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology, 39*(2), 133–146. https://doi.org/10.1017/S0048577202001373

Felser, C., & Cunnings, I. (2012). Processing reflexives in a second language: The timing of structural and discourse-level constraints. *Applied Psycholinguistics, 33*(3), 571–603. https://doi.org/10.1017/S0142716411000488

Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics, 24*(3), 453–489. https://doi.org/10.1017/S0142716403000237

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology, 47*, 164–203. https://doi.org/10.1037/xlm0000519

Ferreira, F., & Chantavarin, S. (2018). Integration and prediction in language processing: A synthesis of old and new. *Current Directions in Psychological Science, 27*(6), 443–448. https://doi.org/10.1177/0963721418794491

Ferreira, F., & Patson, N. D. (2007). The 'Good Enough' Approach to Language Comprehension. *Language and Linguistics Compass, 1*, 71–83. https://doi.org/10.1111/j.1749-818x.2007.00007.x

Fodor, J. (1983). *The modularity of mind*. MIT Press.

Folk, J. R., & Morris, R. K. (2003). Effects of syntactic category assignment on lexical ambiguity resolution in reading: An eye movement analysis. *Memory and Cognition, 31*(1), 87–99. https://doi.org/10.3758/BF03196085

Foucart, A., Garcia, X., Ayguasanosa, M., Thierry, G., Martin, C., & Costa, A. (2015). Does the speaker matter? Online processing of semantic and pragmatic information in L2 speech comprehension. *Neuropsychologia, 75*, 291–303. https://doi.org/10.1016/j.neuropsychologia.2015.06.027

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 601–681). Erlbaum.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition, 6*(4), 291–325. https://doi.org/10.1016/0010-0277(78)90002-1

Frazier, L., & Rayner, K. (1982). *Making and Correcting Errors during Sentence Comprehension: Eye Movements in the Analysis of Structurally Ambiguous Sentences. 210*, 178–210.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*(2), 78–84. https://doi.org/10.1016/S1364-6613(00)01839-8

Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiological Reviews, 91*(4), 1357–1392. https://doi.org/10.1152/physrev.00006.2011

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, *37*(1), 58–93. https://doi.org/10.1006/jmla.1997.2512

Gerth, S., Otto, C., Felser, C., & Nam, Y. (2017). Strength of garden-path effects in native and non-native speakers' processing of object–subject ambiguities. *International Journal of Bilingualism*, *21*(2), 125–144. https://doi.org/10.1177/1367006915604401

Gibson, E. (1992). On the Adequacy of the Competition Model. *Language*, *68*(4), 812–830. https://doi.org/https://www.jstor.org/stable/416855

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76. https://doi.org/10.1016/S0010-0277(98)00034-1

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning Memory and Cognition*, *27*(6), 1411–1423. https://doi.org/10.1037/0278-7393.27.6.1411

Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language*, *51*(1), 97–114. https://doi.org/10.1016/j.jml.2004.02.003

Grüter, T., Lau, E., & Ling, W. (2020). How classifiers facilitate predictive processing in L1 and L2 Chinese: the role of semantic and grammatical cues. *Language, Cognition and Neuroscience*, *35*(2), 221–234. https://doi.org/10.1080/23273798.2019.1648840

Guerreiro, M. J. S., Adam, J. J., & Van Gerven, P. W. M. (2014). Aging and response interference across sensory modalities. *Psychonomic Bulletin and Review*, *21*(3), 836–842. https://doi.org/10.3758/s13423-013-0554-5

Guerreiro, M. J. S., Anguera, J. A., Mishra, J., Van Gerven, P. W. M., & Gazzaley, A. (2014a). Age-equivalent Top–Down Modulation during Cross-modal Selective Attention. *Journal of Cognitive Neuroscience*, *26*(12), 2827–2839. https://doi.org/10.1162/jocn_a_00685

Guerreiro, M. J. S., Eck, J., Moerel, M., Evers, E. A. T., & Gerven, P. W. M. V. (2015). Top-down modulation of visual and auditory cortical processing in aging. *Behavioural Brain Research*, *278*(2015), 226–234. https://doi.org/10.1016/j.bbr.2014.09.049

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience*, *15*(6), 883–899. https://doi.org/10.1162/089892903322370807

Hahne, A., & Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cognitive Brain Research*, *13*(3), 339–356. https://doi.org/10.1016/S0926-6410(01)00127-6

Hahne, A., & Jescheniak, J. D. (2001). What' s left if the Jabberwock gets the semantics? An ERP investigation into semantic and syntactic processes during auditory sentence comprehension. *Cognitive Brain Research*, *11*, 199–212.

Hammarström, H. (2016). Linguistic diversity and language evolution. *Journal of Language Evolution*, *1*(1), 19–29. https://doi.org/10.1093/jole/lzw002

Hardy, S. M., Segaert, K., & Wheeldon, L. (2020). Healthy Aging and Sentence Production: Disrupted Lexical Access in the Context of Intact Syntactic Planning. *Frontiers in Psychology*, *11*(February). https://doi.org/10.3389/fpsyg.2020.00257

Henry, N., Hopp, H., & Jackson, C. N. (2017). Cue additivity and adaptivity in predictive processing. *Language, Cognition and Neuroscience*, *32*(10), 1229–1249. https://doi.org/10.1080/23273798.2017.1327080

Hopp, H. (2014). Working Memory Effects in the L2 Processing of Ambiguous Relative Clauses. *Language Acquisition*, *21*(3), 250–278. https://doi.org/10.1080/10489223.2014.892943

Hopp, H. (2015a). Individual differences in the second language processing of object-subject ambiguities. *Applied Psycholinguistics*, *36*(2), 129–173. https://doi.org/10.1017/S0142716413000180

Hopp, H. (2015b). Semantics and morphosyntax in predictive L2 sentence processing. *IRAL - International Review of Applied Linguistics in Language Teaching*, *53*(3), 277–306. https://doi.org/10.1515/iral-2015-0014

Hsu, N., Badura, K. L., Newman, D. A., & Speach, M. E. P. (2021). Gender, "masculinity," and "femininity": A meta-analytic review of gender differences in agency and communion. *Psychological Bulletin*, *147*(10), 987–1011. https://doi.org/10.1037/bul0000343

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, *1626*, 118–135. https://doi.org/10.1016/j.brainres.2015.02.014

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

Huettig, F., & McQueen, J. M. (2011). The nature of the visual environment induces implicit biases during language-mediated visual search. *Memory and Cognition*, *39*(6), 1068–1084. https://doi.org/10.3758/s13421-011-0086-z

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. https://doi.org/10.1016/j.actpsy.2010.11.003

Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *Journal of Memory and Language*, *98*(October), 1–11. https://doi.org/10.1016/j.jml.2017.09.002

Jackson, C. N. (2008). Proficiency level and the interaction of lexical and morphosyntactic information during L2 sentence processing. *Language Learning*, *58*(4), 875–909. https://doi.org/10.1111/j.1467-9922.2008.00481.x

Jackson, C. N., & Bobb, S. C. (2009). The processing and comprehension of wh-questions among second language speakers of German. *Applied Psycholinguistics*, *30*(4), 603–636. https://doi.org/10.1017/S014271640999004X

Jacob, G., & Felser, C. (2016). Reanalysis and semantic persistence in native and non-native garden-path recovery. *Quarterly Journal of Experimental Psychology*, *69*(5), 907–925. https://doi.org/10.1080/17470218.2014.984231

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, *194*, 137–194. https://doi.org/https://doi.org/10.1207/s15516709cog2002_1

Kamide, Y., & Kukona, A. (2018). The influence of globally ungrammatical local syntactic constraints on real-time sentence comprehension: Evidence from the visual world paradigm and reading. *Cognitive Science*, *42*(8), 2976–2998. https://doi.org/10.1111/cogs.12694

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, *32*(1), 37–55. https://doi.org/10.1023/A:1021933015362

Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *Quarterly Journal of Experimental Psychology*, *69*(5), 1013–1040. https://doi.org/10.1080/17470218.2015.1053951

Kempe, V., & MacWhinney, B. (1998). The Acquisition of Case Marking By Adult Learners of Russian and German. *Studies in Second Language Acquisition*, *20*(04), 543–587. https://doi.org/10.1017/S0272263198004045

Kempe, V., & MacWhinney, B. (1999). Processing of Morphological and Semantic Cues in Russian and German. *Language and Cognitive Processes*, *14*(2), 129–171. https://doi.org/10.1080/016909699386329

Kemper, S., & Liu, C. J. (2007). Eye movements of young and older adults during reading. *Psychology and Aging*, *22*(1), 84–93. https://doi.org/10.1037/0882-7974.22.1.84

Kemtes, K. A., & Kemper, S. (1997). Younger and older adults' on-line processing of syntactically ambiguous sentences. *Psychology and Aging*, *12*(2), 362–371. https://doi.org/10.1037/0882-7974.12.2.362

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

Kilborn, K. (1992). On-line Integration of Grammatical Information in a Second Language. *Advances in Psychology*, *83*(C), 337–350. https://doi.org/10.1016/S0166-4115(08)61504-6

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, *30*(5), 580–602. https://doi.org/https://doi.org/10.1016/0749-596X(91)90027-H

Knoeferle, P. (2007). Comparing the time course of processing initially ambiguous and unambiguous German Svo/ovs sentences in depicted events. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. F. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 517–533). Elsevier. https://doi.org/10.1016/B978-008044980-7/50026-4

Knoeferle, P. (2016). Characterising visual context effects: active, pervasive, but resource-limited. In P. Knoeferle, P. Pyykkönen-Klauck, & M. W. Crocker (Eds.), *Visually Situated Language Comprehension* (pp. 227–260). John Benjamins Publishing Company. https://doi.org/https://doi.org/10.1075/aicr.93.09kno

Knoeferle, P. (2019). Predicting (variability of) context effects in language comprehension. *Journal of Cultural Cognitive Science*, *3*(2), 141–158. https://doi.org/10.1007/s41809-019-00025-5

Knoeferle, P., & Crocker, M. W. (2005). Incremental effects of mismatch during picture-sentence integration: Evidence from eye-tracking. *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, 1166–1171.

Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science*, *30*(3), 481–529. https://doi.org/10.1207/s15516709cog0000_65

Knoeferle, P., & Crocker, M. W. (2007). The influence of recent scene events on spoken comprehension: Evidence from eye movements. *Journal of Memory and Language*, *57*(4), 519–543. https://doi.org/10.1016/j.jml.2007.01.003

Knoeferle, P., & Guerra, E. (2016). Visually Situated Language Comprehension. *Language and Linguistics Compass*, *10*(2), 66–82. https://doi.org/10.1111/lnc3.12177

Knoeferle, P., & Kreysa, H. (2012). Can speaker gaze modulate syntactic structuring and thematic role assignment during spoken sentence comprehension? *Frontiers in Psychology*, *3*(DEC), 1–15. https://doi.org/10.3389/fpsyg.2012.00538

Knoeferle, P., Carminati, M. N., Abashidze, D., & Essig, K. (2011). Preferential inspection of recent real-world events over future events: Evidence from eye tracking during spoken sentence comprehension. *Frontiers in Psychology*, *2*, 1–12. https://doi.org/10.3389/fpsyg.2011.00376

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, *95*(1), 95–127. https://doi.org/10.1016/j.cognition.2004.03.002

Knoeferle, P., Urbach, T. P., & Kutas, M. (2011). Comprehending how visual context influences incremental sentence processing: Insights from ERPs and picture-sentence verification. *Psychophysiology*, *48*(4), 495–506. https://doi.org/10.1111/j.1469-8986.2010.01080.x

Knoeferle, P., Urbach, T. P., & Kutas, M. (2014). Different mechanisms for role relations versus verb-action congruence effects: Evidence from ERPs in picture-sentence verification. *Acta Psychologica*, *152*, 133–148. https://doi.org/10.1016/j.actpsy.2014.08.004

Kreysa, H., Nunnemann, E., & Knoeferle, P. (2018). Distinct effects of different visual cues on sentence comprehension and later recall: The case of speaker gaze versus depicted actions. *Acta Psychologica*, *188*(February 2017), 220–229. https://doi.org/10.1016/j.actpsy.2018.05.001

Kroeger, P. R. (2012). Filler–gap dependencies and relativization. In *Analyzing syntax: A lexical-functional approach* (pp. 165–191). Cambridge University Press. https://doi.org/10.1017/CBO9780511801693

Kukona, A., Cho, P. W., Magnuson, J. S., & Tabor, W. (2014). Lexical interference effects in sentence processing: Evidence from the visual world paradigm and self-organizing models. *Journal of Experimental Psychology: Learning Memory and Cognition*, *40*(2), 326–347. https://doi.org/10.1037/a0034903

Kukona, A., Fang, S. Y., Aicher, K. A., Chen, H., & Magnuson, J. S. (2011). The time course of anticipatory constraint integration. *Cognition*, *119*(1), 23–42. https://doi.org/10.1016/j.cognition.2010.12.002

Kukona, A., Gaziano, O., Bisson, M.-J., & Jordan, A. (2022). Vocabulary knowledge predicts individual differences in the integration of visual and linguistic constraints. *Language, Cognition and Neuroscience*, *37*(6), 750–765. https://doi.org/10.1080/23273798.2021.2020854

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*(1), 23–49. https://doi.org/10.1016/j.brainres.2006.12.063

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

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, *17*, 117–129.

Lardiere, D. (1998). Case and Tense in the 'fossilized' steady state. *Second Language Research*, *14*(1), 1–26. https://doi.org/10.1191/026765898674105303

Lau, E., & Tanaka, N. (2021). The subject advantage in relative clauses: A review. *Glossa*, *6*(1). https://doi.org/10.5334/GJGL.1343

Lau, E., Polinsky, M., Socolof, M., & Clarke, N. (2022). On the representation of morphosyntactic predictions: ERP evidence from Georgian. In *Issues in kartvelian studies*. Vernon Press.

Leal, T., Slabakova, R., & Farmer, T. A. (2017). The fine-tuning of linguistic expectations over the course of L2 learning. *Studies in Second Language Acquisition*, *39*(3), 493–525. https://doi.org/10.1017/S0272263116000164

Lee, J., & Witzel, J. (2022). Plausibility and structural reanalysis in L1 and L2 sentence comprehension. In *Quarterly Journal of Experimental Psychology* (Issue May). https://doi.org/10.1177/17470218221092400

Lenth, R. V. (2022). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. https://cran.r-project.org/package=emmeans

Levy, R., Fedorenko, E., & Gibson, E. (2013). The syntactic complexity of Russian relative clauses. *Journal of Memory and Language*, *69*(4), 461–495. https://doi.org/10.1016/j.jml.2012.10.005

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

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin and Review*, *21*(4), 861–883. https://doi.org/10.3758/s13423-013-0565-2

Lüdecke, D. (2018). *sjPlot: Data Visualization for Statistics in Social Science* (R package version 2.8.11). https://cran.r-project.org/package=sjPlot

Luo, L., Hendriks, T., & Craik, F. I. M. (2007). *Age Differences in Recollection : Three Patterns of Enhanced Encoding*. *22*(2), 269–280. https://doi.org/10.1037/0882-7974.22.2.269

MacWhinney, B. (2002). Extending the Competition Model. In R. R. Heredia & J. Altarriba (Eds.), *Bilingual Sentence Processing* (pp. 31–57). Elsevier Science B. V.

MacWhinney, B. (2005). Second language acquisition and the Competition Model. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in Bilingualism: Psycholinguistic Perspectives* (pp. 113–142). Lawrence Erlbaum Associates.

MacWhinney, B., & Pléh, C. (1997). Double agreement: Role identification in Hungarian. *Language and Cognitive Processes*, *12*(1), 67–102. https://doi.org/10.1080/016909697386916

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior*, 23(2), 127–150. https://doi.org/10.1016/S0022-5371(84)90093-8

Madden, C. J., & Dijkstra, K. (2010). Contextual constraints in situation model construction: An investigation of age and reading span. *Aging, Neuropsychology, and Cognition*, 17(1), 19–34. https://doi.org/10.1080/13825580902927604

Maguinness, C., Setti, A., Burke, K. E., Kenny, R. A., & Newell, F. N. (2011). The effect of combined sensory and semantic components on audio-visual speech perception in older adults. *Frontiers in Aging Neuroscience*, 3(DEC), 1–9. https://doi.org/10.3389/fnagi.2011.00019

Mahr, T. (2020). *Rprime: Functions for working with 'eprime' text files*. https://CRAN.R-project.org/package=rprime

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech Language and Hearing Research*, 50(4), 940.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, 8(1), 1–71. https://doi.org/10.1016/0010-0277(80)90015-3

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, 88(5), 375–407. https://doi.org/10.1037/0033-295X.88.5.375

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, 114(2), 159–188. https://doi.org/10.1037/0096-3445.114.2.159

McRae, K., & Matsuki, K. (2013). Constraint-based models of sentence processing. In R. P. G. van Gompel (Ed.), *Current issues in the psychology of language. Sentence processing* (1st ed., pp. 51–77). Psychology Press.

Meng, M., & Bader, M. (2021). Does comprehension (sometimes) go wrong for noncanonical sentences? *Quarterly Journal of Experimental Psychology*, 74(1), 1–28. https://doi.org/10.1177/1747021820947940

Miklashevsky, A., Reifegerste, J., García, A. M., Pulvermüller, F., Balota, D. A., & Ullman, M. T. (under review). *The processing of action words is impacted by aging, but the task also matters*.

Milisavljevic, A., Abate, F., Le Bras, T., Gosselin, B., Mancas, M., & Doré-Mazars, K. (2021). Similarities and Differences Between Eye and Mouse Dynamics During Web Pages Exploration. *Frontiers in Psychology*, 12(March), 1–16. https://doi.org/10.3389/fpsyg.2021.554595

Mirman, D., & Magnuson, J. S. (2009). Dynamics of activation of semantically similar concepts during spoken word recognition. *Memory and Cognition*, 37(7), 1026–1039. https://doi.org/10.3758/MC.37.7.1026

Mitsugi, S., & MacWhinney, B. (2016). The use of case marking for predictive processing in second language Japanese. *Bilingualism*, *19*(1), 19–35. https://doi.org/10.1017/S1366728914000881

Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language Learning*, *60*(1), 154–193. https://doi.org/10.1111/j.1467-9922.2009.00554.x

Münster, K., & Knoeferle, P. (2017). Situated language processing across the lifespan: A review. *International Journal of English Linguistics*, *7*(1), 1. https://doi.org/10.5539/ijel.v7n1p1

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for Mild Cognitive Impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Nyberg, L., Lövdén, M., Riklund, K., Lindenberger, U., & Bäckman, L. (2012). Memory aging and brain maintenance. *Trends in Cognitive Sciences*, *16*(5), 292–305. https://doi.org/10.1016/j.tics.2012.04.005

Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP Study of Second Language Learning after Childhood: Effects of Proficiency. *Journal of Cognitive Neuroscience*, *17*(8), 1212–1228. https://doi.org/10.1162/0898929055002436

Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language*, *59*(3), 494–522. https://doi.org/10.1006/brln.1997.1793

Pakulak, E., & Neville, H. J. (2010). Proficiency differences in syntactic processing of monolingual native speakers indexed by event-related potentials. *Journal of Cognitive Neuroscience*, *22*(12), 2728–2744. https://doi.org/10.1162/jocn.2009.21393

Pakulak, E., & Neville, H. J. (2011). Maturational constraints on the recruitment of early processes for syntactic processing. *Journal of Cognitive Neuroscience*, *23*(10), 2752–2765. https://doi.org/10.1162/jocn.2010.21586

Peelle, J. E., Troiani, V., Wingfield, A., & Grossman, M. (2010). Neural processing during older adults' comprehension of spoken sentences: Age differences in resource allocation and connectivity. *Cerebral Cortex*, *20*(4), 773–782. https://doi.org/10.1093/cercor/bhp142

Pelletier, A., Bernard, C., Dilharreguy, B., Helmer, C., Le Goff, M., Chanraud, S., … Catheline, G. (2017). Patterns of brain atrophy associated with episodic memory and semantic fluency decline in aging. *Aging*, *9*(5), 1470–1470. https://doi.org/10.18632/aging.101241

Poulisse, C., Wheeldon, L., & Segaert, K. (2019). Evidence against preserved syntactic comprehension in healthy aging. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(12), 2290–2308. https://doi.org/10.1037/xlm0000707

Psychology Software Tools, Inc. (2016). *E-Prime 3.0*. Pittsburgh, PA

Psychology Software Tools, Inc. (2020). *E-PrimeGO 1.0*. Pittsburgh, PA

Puebla, C., & Felser, C. (2022). Discourse prominence and antecedent mis-retrieval during native and non-native pronoun resolution. *Discours*, *29*. https://doi.org/10.4000/discours.11720

Pulvermüller, F., Härle, M., & Hummel, F. (2001). Walking or talking?: Behavioral and neurophysiological correlates of action verb processing. *Brain and Language*, *78*(2), 143–168. https://doi.org/10.1006/brln.2000.2390

Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience*, *21*(3), 793–797. https://doi.org/10.1111/j.1460-9568.2005.03900.x

Pulvermüller, F., Kherif, F., Hauk, O., Mohr, B., & Nimmo-Smith, I. (2009). Distributed cell assemblies for general lexical and category-specific semantic processing as revealed by fMRI cluster analysis. *Human Brain Mapping*, *30*(12), 3837–3850. https://doi.org/10.1002/hbm.20811

Pulvermüller, F., Shtyrov, Y., & Hauk, O. (2009). Understanding in an instant: Neurophysiological evidence for mechanistic language circuits in the brain. *Brain and Language*, *110*(2), 81–94. https://doi.org/10.1016/j.bandl.2008.12.001

Pulvermüller, F., Shtyrov, Y., & Ilmoniemi, R. (2005). Brain signatures of meaning access in action word recognition. *Journal of Cognitive Neuroscience*, *17*(6), 884–892. https://doi.org/10.1162/0898929054021111

R Core Team (2022). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. https://www.R-project.org/

Radvansky, G. A. (1998). The organization of information retrieved from situation models. *Psychonomic Bulletin & Review*, *5*(2), 283–289. https://doi.org/10.3758/BF03212952

Radvansky, G. A., Copeland, D. E., & Zwaan, R. A. (2003). Aging and functional spatial relations in comprehension and memory. *Psychology and Aging*, *18*(1), 161–165. https://doi.org/10.1037/0882-7974.18.1.161

Radvansky, G. A., Zwaan, R. A., Curiel, J. M., & Copeland, D. E. (2001). Situation models and aging. *Psychology and Aging*, *16*(1), 145–160. https://doi.org/10.1037/0882-7974.16.1.145

Rangel Ferrari, L., dos Santos Rodrigues, E., & Correa Mograbi, D. (2019). Garden-path sentences and executive functions in normal aging. *ExLing 2019: Proceedings of 10th International Conference of Experimental Linguistics, 10*(September), 93–96. https://doi.org/10.36505/exling-2019/10/0023/000385

Rankin, T. (2014). Word order and case in the comprehension of L2 German by L1 English speakers. *EUROSLA Yearbook*, *14*, 201–224. https://doi.org/10.1075/eurosla.14.08ran

Reifegerste, J. (2021). The effects of aging on bilingual language: What changes, what doesn't, and why. *Bilingualism*, *24*(1), 1–17. https://doi.org/10.1017/S1366728920000413

Reifegerste, J., & Felser, C. (2017). Effects of aging on interference during pronoun resolution. *Journal of Speech, Language, and Hearing Research*, *60*(12), 3573–3589. https://doi.org/10.1044/2017_JSLHR-L-17-0183

Reifegerste, J., Hauer, F., & Felser, C. (2017). Agreement processing and attraction errors in aging: evidence from subject-verb agreement in German. *Aging, Neuropsychology, and Cognition*, *24*(6), 672–702. https://doi.org/10.1080/13825585.2016.1251550

Reifegerste, J., Jarvis, R., & Felser, C. (2020). Effects of chronological age on native and nonnative sentence processing: Evidence from subject-verb agreement in German. *Journal of Memory and Language*, *111*(January), 104083. https://doi.org/10.1016/j.jml.2019.104083

Reifegerste, J., Meyer, A. S., Zwitserlood, P., & Ullman, M. T. (2021). Aging affects steaks more than knives: Evidence that the processing of words related to motor skills is relatively spared in aging. *Brain and Language*, *218*(December 2020), 104941. https://doi.org/10.1016/j.bandl.2021.104941

Reifegerste, J., Russell, L. E., Balota, D. A., Lipscomb, D., Luta, G., Meinzer, M., Rugg, M. D., Shattuck, K., Turkeltaub, P. E., VanMeter, A. S., Veríssimo, J., & Ullman, M. T. (2022). What's that word again? The contribution of the hippocampus to word-finding declines in aging. Poster presentation at the 2022 Cognitive Aging Conference, Atlanta GA, USA.

Rey-Mermet, A., & Gade, M. (2018). Inhibition in aging: What is preserved? What declines? A meta-analysis. *Psychonomic Bulletin & Review*, *25*(5), 1695–1716. https://doi.org/10.3758/s13423-017-1384-7

Roberts, L., & Felser, C. (2011). Plausibility and recovery from garden paths in second language sentence processing. *Applied Psycholinguistics*, *32*(2), 299–331. https://doi.org/10.1017/S0142716410000421

Rosenblum, L. D. (2005). Primacy of multimodal speech perception. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 51–78). Blackwell.

Rosenblum, L. D. (2008). Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science*, *17*(6), 405–409. https://doi.org/10.1111/j.1467-8721.2008.00615.x

Rossi, S., Gugler, M. F., Friederici, A. D., & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *Journal of Cognitive Neuroscience*, *18*(12), 2030–2048.

Samu, D., Campbell, K. L., Tsvetanov, K. A., Shafto, M. A., Brayne, C., Bullmore, E. T., … Tyler, L. K. (2017). Preserved cognitive functions with age are determined by domain-dependent shifts in network responsivity. *Nature Communications*, *8*(May). https://doi.org/10.1038/ncomms14743

Sánchez-Izquierdo, M., & Fernández-Ballesteros, R. (2021). Cognition in healthy aging. *International Journal of Environmental Research and Public Health*, *18*(3), 1–30. https://doi.org/10.3390/ijerph18030962

Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences*, *6*(9), 382–386. https://doi.org/10.1016/S1364-6613(02)01958-7

Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, *56*(6), 1–8. https://doi.org/10.1111/psyp.13335

Sato, M., & Felser, C. (2010). Sensitivity to Morphosyntactic Violations in English as a Second Language. *Second Language*, *9*, 101–118. https://doi.org/10.11431/secondlanguage.9.0_101

Schlenter, J., & Felser, C. (2021). Second language prediction ability across different linguistic domains. In E. Kaan & T. Grüter (Eds.), *Prediction in second language processing and learning* (pp. 48–68). John Benjamins. https://doi.org/10.1075/bpa.12.03sch

Schlesewsky, M., & Bornkessel, I. (2004). On incremental interpretation: Degrees of meaning accessed during sentence comprehension. *Lingua*, *114*(9-10), 1213–1234. https://doi.org/10.1016/j.lingua.2003.07.006

Schlesewsky, M., Fanselow, G., Kliegl, R., & Krems, J. (2000). The subject preference in the processing of locally ambiguous wh-questions in German. In B. Hemforth & L. Konieczny (Eds.), *German sentence*

processing. Studies in theoretical psycholinguistics (pp. 65–93). Springer. https://doi.org/10.1007/978-94-015-9618-3_3

Schmitter-Edgecombe, M. (2000). Aging and word-finding: A comparison of spontaneous and constrained naming tests. *Archives of Clinical Neuropsychology*, *15*(6), 479–493. https://doi.org/10.1016/S0887-6177(99)00039-6

Schoemann, M., Lüken, M., Grage, T., Kieslich, P. J., & Scherbaum, S. (2019). Validating mouse-tracking: How design factors influence action dynamics in intertemporal decision making. *Behavior Research Methods*, *51*(5), 2356–2377. https://doi.org/10.3758/s13428-018-1179-4

Schoemann, M., O'Hora, D., Dale, R., & Scherbaum, S. (2021). Using mouse cursor tracking to investigate online cognition: Preserving methodological ingenuity while moving toward reproducible science. *Psychonomic Bulletin and Review*, *28*(3), 766–787. https://doi.org/10.3758/s13423-020-01851-3

Schriefers, H., Friederici, A. D., & Kühn, K. (1995). The processing of locally ambiguous relative clauses in German. *Journal of Memory and Language*, *34*(4), 499–520. https://doi.org/10.1006/jmla.1995.1023

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

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*(4), 523–568. https://doi.org/10.1037//0033-295x.96.4.523

Service, E., Maury, S., & Salmelin, R. (2007). Localization of syntactic and semantic brain responses using magnetoencephalography. *Journal of Cognitive Neuroscience*, *19*(7), 1193–1205.

Sevcenco, A., Avram, L., & Stoicescu, I. (2013). Subject and direct object relative clause production in child Romanian. In L. Avram & A. Sevcenco (Eds.), *Topics in Language Acquisition and Language Learning in a Romanian Context* (pp. 51–85). Editura Universității din București.

Shake, M. C., & Stine-Morrow, E. A. L. (2016). Language and aging. *The Curated Reference Collection in Neuroscience and Biobehavioral Psychology*, (September 2015), 337–342. https://doi.org/10.1016/B978-0-12-809324-5.01889-7

Spivey, M. J., & Huette, S. (2016). Towards a situated view of language. In P. Knoeferle, P. Pyykkönen-Klauck, & M. W. Crocker (Eds.), *Visually situated language comprehension* (pp. 1–30). John Benjamins Publishing Company. https://doi.org/10.1075/aicr.93.01spi

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, *45*(4), 447–481. https://doi.org/10.1016/S0010-0285(02)00503-0

Spivey-Knowlton, M. J., Trueswell, J. C., & Tanenhaus, M. K. (1993). Context effects in syntactic ambiguity resolution: discourse and semantic influences in parsing reduced relative clauses. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Expérimentale*, *47*(2), 276–309. https://doi.org/10.1037/h0078826

Stine-Morrow, E. A. L. (2007). The Dumbledore Hypothesis of Cognitive Aging. *Current Directions in Psychological Science*, *16*(6), 295–299. https://doi.org/10.1111/j.1467-8721.2007.00524.x

Stine-Morrow, E. A. L., & Payne, B. R. (2016). Age Differences in Language Segmentation. *Experimental Aging Research*, *42*(1), 107–125. https://doi.org/10.1080/0361073X.2016.1108751

Stine-Morrow, E. A. L., Loveless, M. K., & Soederberg, L. M. (1996). Resource allocation in on-line reading by younger and older adults. *Psychology and Aging*, *11*(3), 475–486. https://doi.org/10.1037/0882-7974.11.3.475

Stine-Morrow, E. A. L., Ryan, S., & Leonard, J. S. (2000). Age differences in on-line syntactic processing. *Experimental Aging Research*, *26*(4), 315–322. https://doi.org/10.1080/036107300750015714

Stone, K., Lago, S., & Schad, D. J. (2021). Divergence point analyses of visual world data: Applications to bilingual research. *Bilingualism*, *24*(5), 833–841. https://doi.org/10.1017/S1366728920000607

Swets, B., Desmet, T., Clifton, C., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory and Cognition*, *36*(1), 201–216. https://doi.org/10.3758/MC.36.1.201

Taler, V., Klepousniotou, E., & Phillips, N. A. (2009). Comprehension of lexical ambiguity in healthy aging, mild cognitive impairment, and mild Alzheimer's disease. *Neuropsychologia*, *47*(5), 1332–1343. https://doi.org/10.1016/j.neuropsychologia.2009.01.028

Tanenhaus, M. K., Carroll, J. M., & Bever, T. G. (1976). Sentence-picture verification models as theories of sentence comprehension: A critique of Carpenter and Just. *Psychological Review*, *83*(4), 310–317. https://doi.org/10.1037/0033-295X.83.4.310

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

Tanner, D., Inoue, K., & Osterhout, L. (2014). Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism*, *17*(2), 277–293. https://doi.org/10.1017/S1366728913000370

Tanner, D., McLaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism*, *16*(2), 367–382. https://doi.org/10.1017/S1366728912000302

Taraban, R., & McClelland, J. L. (1988). Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectations. *Journal of Memory and Language*, *27*(6), 597–632. https://doi.org/10.1016/0749-596X(88)90011-3

The GIMP Development Team. (2019). GIMP. Retrieved from https://www.gimp.org

Thomann, A. E., Goettel, N., Monsch, R. J., Berres, M., Jahn, T., Steiner, L. A., & Monsch, A. U. (2018). The Montreal Cognitive Assessment: Normative data from a German-speaking cohort and comparison with international normative samples. *Journal of Alzheimer's Disease*, *64*(2), 643–655. https://doi.org/10.3233/JAD-180080

Thomas, A. K., Dave, J. B., & Bonura, B. M. (2019). Theoretical Perspectives on Cognitive Aging. In C. L. Armstrong & L. A. Morrow (Eds.), *Handbook of Medical Neuropsychology* (2nd ed.). Springer Nature Switzerland AG 2019. https://doi.org/10.1007/978-3-030-14895-9

Tomasello, M., Hare, B., Lehmann, H., & Call, J. (2007). Reliance on head versus eyes in the gaze following of great apes and human infants: the cooperative eye hypothesis. *Journal of Human Evolution*, *52*(3), 314–320. https://doi.org/10.1016/j.jhevol.2006.10.001

Trabasso, T., Rollins, H., & Shaughnessy, E. (1971). Storage and verification stages in processing concepts. *Cognitive Psychology*, *2*(3), 239–289. https://doi.org/10.1016/0010-0285(71)90014-4

Traxler, M. J. (2011). Parsing. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(4), 353–364. https://doi.org/10.1002/wcs.112

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*(3), 285–318. https://doi.org/10.1006/jmla.1994.1014

Tse, C., & Altarriba, J. (2014). *The relationship between language proficiency and attentional control in Cantonese-English bilingual children : evidence from Simon , Simon switching , and working memory tasks*. *5*(September), 1–18. https://doi.org/10.3389/fpsyg.2014.00954

Tun, P. A., McCoy, S., & Wingfield, A. (2009). Aging, Hearing Acuity, and the Attentional Costs of Effortful Listening. *Psychology and Aging*, *24*(3), 761–766. https://doi.org/10.1037/a0014802

Tyler, L. K., Shafto, M. A., Randall, B., Wright, P., Marslen-Wilson, W. D., & Stamatakis, E. A. (2010). Preserving syntactic processing across the adult life span: The modulation of the frontotemporal language system in the context of age-related atrophy. *Cerebral Cortex*, *20*(2), 352–364. https://doi.org/10.1093/cercor/bhp105

Underwood, G., Jebbett, L., & Roberts, K. (2004). Inspecting pictures for information to verify a sentence: Eye movements in general encoding and in focused search. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *57*(1), 165–182. https://doi.org/10.1080/02724980343000189

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning Memory and Cognition*, *31*(3), 443–467. https://doi.org/10.1037/0278-7393.31.3.443

van Boxtel, W., & Lawyer, L. (2021). Sentence comprehension in ageing and Alzheimer's disease. *Language and Linguistics Compass*, *15*(6). https://doi.org/10.1111/lnc3.12430

Van Gompel, R. P. G., & Pickering, M. J. (2006). Syntactic parsing. In M. Traxler & M. A. Gernsbacher (Eds.), *The oxford handbook of psycholinguistics* (2nd ed., pp. 455–503). Academic Press. https://doi.org/10.1093/oxfordhb/9780198568971.013.0017

Verhaegen, C., & Poncelet, M. (2013). Changes in Naming and Semantic Abilities With Aging From 50 to 90 years. *Journal of the International Neuropsychological Society*, *19*(2), 119–126. https://doi.org/https://doi.org/10.1017/S1355617712001178

Veríssimo, J., Heyer, V., Jacob, G., & Clahsen, H. (2018). Selective Effects of Age of Acquisition on Morphological Priming: Evidence for a Sensitive Period. *Language Acquisition*, *25*(3), 315–326. https://doi.org/10.1080/10489223.2017.1346104

Veríssimo, J., Verhaeghen, P., Goldman, N., Weinstein, M., & Ullman, M. T. (2022). Evidence that ageing yields improvements as well as declines across attention and executive functions. *Nature Human Behaviour*, *6*(1), 97–110. https://doi.org/10.1038/s41562-021-01169-7

Wannemacher, J. T. (1974). Processing strategies in picture-sentence verification tasks. *Memory & Cognition*, *2*(3), 554–560. https://doi.org/10.3758/BF03196920

Weber, A., Grice, M., & Crocker, M. W. (2006). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition*, *99*, 63–72. https://doi.org/https://doi.org/10.1016/j.cognition.2005.07.001

Wellsby, M., & Pexman, P. M. (2014). Developing embodied cognition: Insights from children's concepts and language processing. *Frontiers in Psychology*, *5*(MAY), 1–10. https://doi.org/10.3389/fpsyg.2014.00506

Wickham, H. (2008). *Elegant Graphics for Data Analysis: ggplot2* (pp. 21–54).

Wickham, H. (2016). *Elegant Graphics for Data Analysis Second Edition* (Vol. 35). Retrieved from http://www.springer.com/series/6991

Wingfield, A., Amichetti, N. M., & Lash, A. (2015). Cognitive aging and hearing acuity: Modeling spoken language comprehension. *Frontiers in Psychology*, *6*(MAY), 1–13. https://doi.org/10.3389/fpsyg.2015.00684

Wingfield, A., Peelle, J. E., & Grossman, M. (2003). Speech Rate and Syntactic Complexity as Multiplicative Factors in Speech Comprehension by Young and Older Adults. *Aging, Neuropsychology, and Cognition*, *10*(4), 310–322. https://doi.org/10.1076/anec.10.4.310.28974

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

Yadav, H., Smith, G., Mertzen, D., Engbert, R., & Vasishth, S. (2022). Individuals differ cross-linguistically in cue weighting: A computational evaluation of cue- based retrieval in sentence processing. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*(44). https://doi.org/https://doi.org/10.1162/opmi_a_00052

Yildirim, I., & Jacobs, R. A. (2013). Transfer of object category knowledge across visual and haptic modalities: Experimental and computational studies. *Cognition*, *126*(2), 135–148. https://doi.org/10.1016/j.cognition.2012.08.005

Yildirim, I., & Jacobs, R. A. (2015). Learning multisensory representations for auditory-visual transfer of sequence category knowledge: a probabilistic language of thought approach. *Psychonomic Bulletin and Review*, *22*(3), 673–686. https://doi.org/10.3758/s13423-014-0734-y

Yoon, J., Campanelli, L., Goral, M., Marton, K., Eichorn, N., & Obler, L. K. (2015). The effect of plausibility on sentence comprehension among older adults and its relation to cognitive functions. *Experimental Aging Research*, *41*(3), 272–302. https://doi.org/10.1080/0361073X.2015.1021646

Zhang, Y., Frassinelli, D., Tuomainen, J., Skipper, J. I., & Vigliocco, G. (2021). More than words: Word predictability, prosody, gesture and mouth movements in natural language comprehension. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1955). https://doi.org/10.1098/rspb.2021.0500

Zhou, H., Rossi, S., Li, J., Liu, H., Chen, R., & Chen, B. (2017). Effects of working memory capacity in processing wh-extractions: eye-movement evidence from Chinese–English bilinguals. *Journal of Research in Reading*, *40*(4), 420–438. https://doi.org/10.1111/1467-9817.12079

Zona, C. I. (under review). The temporal dynamics of incongruence resolution for situated thematic-role prediction. *Journal of Psycholinguistic Research*.

Zona, C. I., & Felser, C. (2023). Integrating morphosyntactic and visual cues in L1 and L2 comprehension. *Languages*, *8*(2), 111. https://doi.org/10.3390/languages8020111

Zona, C. I., & Reifegerste, J. (under review). The weighting of syntactic versus visual-context information during comprehension: A lifespan perspective. *Psychology and Aging*.

Zubin, D. A. (1979). Discourse function of morphology: The focus system in German. In *Discourse and syntax* (pp. 469–504). BRILL. https://doi.org/10.1163/9789004368897_020

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162–185. https://doi.org/10.1037/0033-2909.123.2.162

Declaration of authorship

I hereby certify that the thesis I am submitting is entirely my own original work except where otherwise indicated. I am aware of the University's regulations concerning plagiarism, including those regulations concerning disciplinary actions that may result from plagiarism. Any use of the works by any other author, in any form, is properly acknowledged at their point of use.

_____

Date, Signature

Eigenständigkeitserklärung

Hiermit bestätige ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken (dazu zählen auch Internetquellen) entnommen sind, wurden unter Angabe der Quelle kenntlich gemacht.

_____

(Datum, Unterschrift)