REVIEW

Predicting (variability of) context effects in language comprehension

Pia Knoeferle

Received: 15 February 2019 / Revised: 9 July 2019 / Accepted: 11 July 2019 / Published online: 28 August 2019
© The Author(s) 2019

Abstract Predicting variability in context effects is a timely enterprise considering that psycho- and neurolinguistic research has assessed how language processing depends on the perceived context, the body, and long-term linguistic knowledge of the language user. The current evidence suggests that some context effects may be systematically more robust than others and that language user characteristics are an influential modulator of context-sensitive comprehension. Reviewing psycholinguistic evidence, I argue for constrained contextual variability. Variability in context effects is predicted by characteristics of the language user and world-language relations. But extant findings also suggest generalizability beyond such variation, thus imposing constraint on theoretical prediction of context effects via relative (not absolute) processing preferences.

Keywords Visually-situated language comprehension · Predicting context effects ·

Variability · Systematicity · Language user characteristics

A focus on language and linguistic experience in theories of language comprehension

Psycholinguistic approaches to sentence comprehension have investigated what mechanisms and what representations we must assume to accommodate how comprehenders recover an interpretation from language input in real time. Early theories of real-time comprehension (Frazier and Fodor 1979), for instance, described a ‘sausage machine’ (p. 291) for which tree diagrams with abstract nodes represented words and syntactic relations between them. The authors proposed a two-stage mechanism of language comprehension, with a focus on building syntactic structure (called a ‘parse’). The postulated stages can be viewed as a mechanism for how structure is built (e.g., without permitting influence from world knowledge on initial syntactic analysis). An initial stage assigned lexical and phrasal nodes (in a tree diagram) to approximately six words within a sentence (Frazier and Fodor 1979). A second stage added nodes in the tree to build a complete sentence (see Bornkessel and Schlesewsky 2006; Friederici 2002, for neuroscientific accounts of related hierarchical models). If needed, the initial structure and interpretation could, at a second stage, be revised.

P. Knoeferle (✉)
Institut für deutsche Sprache und Linguistik, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany
e-mail: pia.knoeferle@hu-berlin.de

P. Knoeferle
Einstein Center for Neurosciences Berlin, Berlin, Germany

P. Knoeferle
Berlin School of Mind and Brain, Berlin, Germany

These early parsing accounts did not consider further contextual information such as objects, actions, and events in the world and their potential impact on comprehension. Nor did they focus on a model of the world and typical situations such as ordering food in a restaurant as a context for comprehension. Instead they focused on principles regarding syntactic analysis (Frazier and Fodor 1979; Kimball 1973) that were oriented towards maximizing syntactic simplicity. One such principle, for instance, postulated that a phrase was closed as soon as possible [unless the next parsed node was an “immediate constituent” of the phrase, p. 36 in Kimball (1973)]. This accommodates that comprehenders seem to interpret *they knew the girl* as a simple sentence instead of assuming that it might have a different structure as in *they knew the girl was in the closet* (p. 36).

An alternative view of language processing mechanisms that assumed rapid feedback from higher-order expectations and world knowledge on initial syntactic and semantic processes received support early on from so-called ‘shadowing’ tasks. In these, participants were trained to rapidly repeat (‘shadow’) a speaker’s language production (Marslen-Wilson 1973) at very short lags (254–278 ms for the best shadowers). Participants sometimes made errors in that task; the rationale was that if shadowers’ errors violated the speaker’s preceding syntax or semantics at short lags of shadowing, then one could assume that shadowers are not influenced by higher-level content during ongoing comprehension and production. Crucially, when shadowers made errors, these virtually never conflicted syntactically or semantically with the preceding linguistic context. On the basis of this finding, Marslen-Wilson (1973) argued that ‘higher-order linguistic structure’ (p. 522f. Marslen-Wilson 1973) is available during shadowing (comprehension and production) at very short lags. The shadowing findings contradicted postulates of ‘informational encapsulation’ in the philosophy of language (Fodor 1983, p.71). These had assumed that brain systems dealing with incoming language had no immediate access to high-level expectations or to ongoing visual perception (Coltheart 1999, p. 117). However, animacy (Trueswell et al. 1994), plausibility (Garnsey et al. 1997), thematic fit (e.g., such as a cop being a good agent for arresting a criminal McRae et al. 1998), and object arrangements (Tanenhaus et al. 1995; Spivey et al. 2002) all influenced the resolution

of local structural ambiguity within a few hundred milliseconds. The reach of lexical, and associated world knowledge, as well as of visual perception into the resolution of temporary structural ambiguity boosted accounts of sentence processing that accorded a central role to the lexicon (MacDonald et al. 1994) and supported immediate interaction of syntactic parsing also with visual perception.

While some accounts accorded a central role to the lexicon and lexical information in structural ambiguity resolution, others postulated that parsing decisions at structural ambiguities are modulated not just by lexical information (Mitchell et al. 1995). For instance, in direct object/subject complement ambiguities such as *The athlete realized her goals (dir. obj.)/her shoes were out of reach (subject complement)...*, the verb biases towards the complement analysis. If lexical experience determined the initial parse, then participants should be biased towards a subject complement analysis. However, Pickering and Traxler (1995) observed more frequent regression when the object was implausible (*shoes*) than plausible (*goals*). The authors interpreted this result as suggesting that readers initially adopted an object analysis and that semantic plausibility rapidly modulated that analysis. Mitchell et al. (1995) argued that readers must use coarse-grained rather than fine-grained lexical experience. By coarse-grained they seem to mean the average “usage of the different structural forms over all verbs which share the same ambiguity”, which they expect to reflect a strong bias towards the direct object interpretation (p. 484). This supports early psycholinguistic claims—viz. that parsing principles guide ambiguity resolution—but motivates attachment decisions from experience with language instead of syntactic simplicity. Thus, both coarse-grained and item-specific experience appear to contribute substantially to real-time language comprehension.

The role of language experience also comes to the fore in research that argued comprehenders anticipate upcoming content (Altmann and Kamide 1999; Kutas and Hillyard 1984). Kutas and Hillyard (1984) provided evidence for the view that brain potentials reflected the expectancy of a word and of semantic associations during reading. The amplitude of a negative-going brain potential around 400 ms post stimulus, the so-called N400 (for a review see Kutas and Federmeier 2011) was smaller the more participants expected a sentence-final word (as measured by

cloze probability). In psycholinguistics, Clifton et al. (1984) reported that readers relied on lexical (verb) information to anticipate noun phrases and structure (filler-gap relations). Predicting the expectedness of words was taken up in computational modeling via an information-theoretic notion dubbed ‘surprisal’ Hale (2003). Surprisal captured—via inverse log probabilities—how unexpected incoming information was. From such expectation measures researchers derived predictions for processing difficulty and, relatedly, measures such as reading times and event-related brain potentials (Frank 2013; Frank et al. 2015; Delogu et al. 2017; Demberg and Keller 2008).

The accounts discussed thus far were motivated by accommodating initial structure-building decisions and the incremental or anticipatory nature of language processing informed by linguistic and world knowledge, and visual context. The focus in these theories was further on accommodating comprehension “on average” with little attention to potential individual differences—with the exception of, for instance, studies on language deficits (e.g., Grossman et al. 1992, 1993), comprehension skill (King and Kutas 1995), and cross-linguistic variation, (Bates et al. 1987; MacWhinney et al. 1984). Empirically, object contexts in psycholinguistics were examined as a factor that could rapidly modulate parsing decisions, providing evidence against strictly informationally encapsulated processing (see Chambers et al. 2002, 2004; Tanenhaus et al. 1995).

But fleshing out the mechanistic and representational implications of integrating rich visual context (including representations of objects, action, events, and speakers, their gaze, mimics, and appearance) were not the focus of the dominant psycholinguistic theories of language comprehension from the late 1970s to the late 1990s. In fact, approaches that investigated how object representations gleaned from a picture are reconciled with language (e.g., in picture-sentence verification paradigms and associated formalized models) were criticized as not revealing comprehension (Tanenhaus et al. 1976). In the meantime, however, empirical research has provided more and more evidence that we must not only consider object and event contexts (and their associated representations) for developing theories of real-time language comprehension but also (representations of) the listener and his characteristics such as age, or educational background among others (see Münster and

Knoeferle 2018; Huettig et al. 2018; Mishra et al. 2012).

Below I first discuss (representational and mechanistic) assumptions of context-focused theoretical approaches to comprehension. From these theoretical approaches, we can identify key processes and associated mechanisms in comprehension, viz. interpretation and structuring, attentional grounding in context (and associated representations), and verification of the interpretation against the context (ual representations). I discuss variation in these key processes and specifically in contextual grounding as a function of comprehender characteristics (“[Variation in expectation-based comprehension and in context effects by comprehender characteristics](#)”) and world-language relations (“[Differentiating grounding: do different world-language relations elicit distinct context effects?](#)”). From that discussion, I derive three principles for predicting variability in context effects (“[Conclusions: predicting \(visual\) context effects](#)”).

Context-focused approaches to comprehension and associated representational assumptions

Early contextualized approaches to language processing have specified mechanistic and representational assumptions but these have failed to influence early models of parsing and linguistic knowledge (Frazier and Fodor 1979; MacDonald et al. 1994; Mitchell et al. 1995). These contextualized approaches can be grouped in approaches that focus on (1) the representations and mechanisms (verification, semantic activation, and listening-looking interaction) of language processing in a physical or object-/event context, and (2) a mental model of a (discourse) context, obtained via inferential processing, drawing on script and event/body knowledge.

Verification, lexical-semantic activation, and listening-looking interaction

To examine language comprehension in object contexts, one paradigm had participants verify a sentence in the context of a picture (e.g., Clark and Chase 1972; Gough 1965). Incongruence (vs. congruence) of the sentence against the picture slowed verification latencies. The latter finding was accommodated in the Constituent-Comparison Model (Carpenter and Just

1975) of sentence comprehension via a serial comparison of picture-based and sentence-based representations (e.g., *The dots are red*, tracking the affirmative nature of the sentence and the picture representations: [AFFIRMATIVE, (RED, DOTS)], Picture of red dots: (RED, DOTS)). Corresponding constituents were retrieved and compared, pair by pair. The ordering of these comparisons was determined by the structure of the representations and inner propositions were compared first. The number of retrieve-and-compare steps determined the predicted verification latencies. The model was criticized (Tanenhaus et al. 1976) as not describing the processes at work while participants derive sentence representations but rather post-perceptual processes of verifying a sentence against an already-perceived picture.

That criticism fell short, however (Knoeferle et al. 2011) since incongruence in picture and sentence rapidly modulated comprehension also at the word that mismatched the visual context (Knoeferle and Crocker 2005) and not only at the point when participants gave the verification response via button press, post-sentence. For instance, when participants verified whether a word (e.g., *red* in *Touched the small blue circle and the large red square*) matched or mismatched the color of a depicted object touched by a hand (e.g., a circle bearing the label 'blue'), participants' event-related brain potentials (mean amplitude negativities) from around 200 ms and peaking between approximately 450 and 750 ms increased to mismatches (vs. matches) (D'Arcy and Connolly 1999).

On the basis of these and other results, it has been argued (Knoeferle et al. 2011) that verification of language against context seems to be a fundamental part of language processing in context. However, despite its occasional use in language comprehension research (Goolkasian 1996; Singer 2006; Underwood et al. 2004) insights from the verification task have had minimal impact on psycholinguistic theories of online sentence comprehension, but see Knoeferle et al. (2014).

Verification is one important mechanism; but further, inferential and semantic interpretation processes are necessary to accommodate language comprehension in context. Indeed, empirical research even in the 1970s had begun to examine how words and pictures are processed to obtain insight into their meaning representation. In one study by Potter and Faulconier (1975), participants saw words or object

drawings, and read the word or named the object in a first condition pair. In further conditions, the experimenter named a category and the participant responded ('yes'/'no') depending on whether the next word/object drawing belonged to that category. For object drawings (vs. words), naming latencies were longer but categorization latencies were shorter. Drawings were also categorized faster than they were named. The conclusion drawn from this finding was that the category knowledge of an object is linked to an abstract object concept (instead of its name/appearance), for recent related evidence in the case of object color see, Amsel et al. (2014) and Connell (2007).

Another line of research assessed representational issues via semantic 'priming'. In priming, stimuli are presented in pairs (a 'prime' stimulus followed by a 'target' stimulus), and these were either semantically related or unrelated in Sperber et al. (1979). Semantic relatedness facilitated naming/reading of pictures and words respectively and interacted with stimulus quality (in vs. out of focus photographs). Priming occurred whether prime and target were both words, both pictures, or mixed (e.g., picture–word), a finding that was interpreted as pictures and words accessing semantic information via a common conceptual store. Priming was, however, more pronounced in picture–picture than word–word or mixed pairs, a finding that Sperber et al. (1979) traced to overlap in pictorial representations of objects from the same semantic category.

Semantic priming is active also as a sentence-level comprehension mechanism in pictorial contexts (Ganis et al. 1996), and could be viewed as related to verification. Ganis et al. (1996) had participants read sentences that were semantically constraining and in which the final word (or picture) was in agreement with the sentence or mismatched (e.g., *The old man lay on the grass and lit his pipe/carrot*). From the findings of this study (earlier peaking and topographically distinct N400 effects to pictures vs. words), the authors concluded that pictures and words had at least partially distinct representations. Beyond representational insights regarding words and objects, priming research has impacted psycholinguistic research and theories, for instance, in production (Bock 1986), but also in comprehension (Nichol and Pickering 1993), though often with linguistic context as prime and target, see (Pickering and Branigan 1999; Pickering and Garrod 2004).

By contrast other, context-focused psycholinguistic research on language processing failed to influence psycholinguistic thinking of the 1970s and 80s (Cooper 1974; Just and Carpenter 1976). Cooper (1974) recorded eye movements to object drawings during spoken discourse comprehension and interpreted them as reflecting comprehension processes (p. 104). The analysis of the collected gaze record contributed three important findings. First, Cooper, discovered that listeners inspected referents upon their mention (e.g., a depiction of a zebra upon hearing *zebra*), suggesting *rapid interaction of auditory comprehension with visual attention*. Moreover, the analyses of Cooper's eye-tracking data provided evidence for the *role of expectations in comprehension*: Listeners inspected, for instance, the depiction of a zebra and of a lion after they had heard *While on a photographic safari in Africa*. Neither the zebra nor the lion had been mentioned, suggesting that the linguistic context and mention of a safari and of Africa elicited expectations—via world knowledge—of as-yet-unmentioned referents. Finally, Cooper (1974) observed that picture-language relations such as between a zebra and the referential expression *zebra* elicited a much higher number of concurrent fixations than other relations such as between the modifier *photographic* and the depiction of a camera). This latter finding can be viewed as reflecting a *prominent status of referential relations* in language processing¹.

Cooper recognized the potential of the eye-tracking method: “Because the eye-movement response system in the presence of ongoing heard language is at times characterized by a high degree of linguistic sensitivity and small latencies (including a built-in anticipatory characteristic), the present technique of correlating the visual selection of appropriate targets with concurrently heard words could potentially be applied to study in great detail the manner in which people interpret and process spoken language in the context of their contemporary visual field.” (p. 106). He also recommended the method for examining referential ambiguity resolution, speech perception, and memory. Relevant ensuing research has indeed followed these suggestions, reporting, among others, effects of object

size, contrast, and of action affordances on the resolution of referential and structural ambiguity (Chambers et al. 2004; Sedivy et al. 1999; Tanenhaus et al. 1995).

The inferential construction of mental models and grounding via bodily schemas

While verification mechanisms, semantic associations, and listening-looking interactions are important for comprehension in context, so are arguably further inferential processes and models of the world. One account by Johnson-Laird (1981) included both of these. It assumed that utterances are conveyed to representations and these can (but need not always) serve as the basis for the inferential construction of a mental model of world. The latter process enables “one person to have another's experience of the world by proxy: instead of direct apprehension of a state of affairs, the listener constructs a model of them based on a speaker's remarks” (1981, p. 139). Utterance interpretation in this proposal depended on script knowledge, viz. of normal event sequences such as lunching at a restaurant (e.g., Schank and Abelson 1975). Understanding non-stereotypical events (e.g., Kafka's *Trial*, Johnson-Laird 1981, p. 154), depended on inferences, on coherence and on plausibility, with recursion as a mechanism for manipulating the world model. However, much like the verification research, the influence of situation-model approaches on theories of sentence comprehension in the 80s and early 90s was negligible. However, they have influenced empirical research and theory formation since, as discussed in more detail below (for a review of embodiment and situation models see Meteyard et al. 2012; Zwaan 2016).

Another line of research that brings grammatical representations in close proximity of real-world perception and action are some versions of construction grammar ('CG'). Central work in CG has assessed how constructions generalize (Goldberg 1995, 2006; Langacker 1987) but a few approaches have related semantic structure to visual or motor representations, for instance, Fluid Construction Grammar, (Trijp et al. 2012), Embodied Construction Grammar (ECG), and Template Construction Grammar, (Arbib and Lee 2008). As one example, ECG assumes embodied cognitive 'schemas', i.e., representations derived from perceptual and motor experience (Bergen and Chang

¹ Note, that Cooper (1974) collapsed referential and superordinate-token relations (e.g., *forrest* eliciting looks to a tree) within one category; however, most relations in that category were referential.

2005). ECG focuses on the image schemas of Lakoff and Johnson (1980) and motor schemas (X-schemas). These representations bring language representations in contact with the bodily-experienced world.

ECG assumes that semantic representations are grounded in comprehenders' perceptual and motor systems. Language comprehension implicates both analysis (what constructions must be built) and associated mental simulation (e.g., of a tossing action). ECG envisages incrementality (Bergen and Chang 2005) but does not specify the activation of constructions, word-by-word (but see Bryant (2008) on reading-time data). To interpret *Mary tossed me a bottle* in ECG, reference from *Mary* to the referent *Mary* is resolved via a referent scheme and *tossed* unlocks a predicate schema (the subject codes the cause of the transfer, the first post-verbal object the recipient, and the second post-verbal object the transferred object), as well as a transfer schema (specifying an agent: *Mary*; a recipient: *me*; and a theme: the bottle). These schemas are integrated into the interpretation, evoking an active-ditransitive schema with context verifying its constraints. The verb *toss* activates a caused-motion schema and a fly schema, specifying low force. X-schemas capture sequential action elements. But it remains open when an X-schema for tossing is activated (e.g., at *tossed*, or even earlier at *Mary* depending on whether *Mary* is inactive or tossing already).

In summary, by the early 2000s, it was clear that key processes in context-sensitive language comprehension include minimally the incremental interpretation of language, rapidly informed by linguistic and world knowledge; its reconciliation with the world ('grounding'), mediated via (visual) attention to (depicted) objects in the context; and verification of the interpretation in relation to objects and events. Representationally, it had become clear that mental representations for grounding must include objects and their properties, action events and experience-based event knowledge (e.g. object affordances, scripts, bodily schemas). In addition, research in the cognitive sciences has examined, for instance, how authorship (social speaker identity) can constrain the interpretation of words (Fitneva and Spivey 2005), or how linguistic and para-linguistic information contribute to establishing reference and common ground between a speaker and a listener (e.g., Clark and Brennan 1991). Thus, by the early 2000s, empirically,

it began to be clear that social aspects and grounding, too, play an important role for language processing. Below I first outline to what extent the contextualized approaches have impacted psycholinguistic theorizing; I enumerate key processes in situated language comprehension and discuss their variability ("Three key processing steps and associated variability"). Then I focus on one process—grounding and context effects—and discuss comprehender-based variation in context effects ("Variation in expectation-based comprehension and in context effects by comprehender characteristics"). Finally, I discuss distinct world-language relations as one source of variability in context effects ("Differentiating grounding: do different world-language relations elicit distinct context effects?") and derive from this discussion predictions of context effects ("Conclusions: predicting (visual) context effects").

Towards predicting context effects

Given the insights from (only a small sample of) context-focused approaches to language comprehension, one may ask to what extent these approaches were integrated with, or have influenced, earlier theories of sentence comprehension (see "A focus on language and linguistic experience in theories of language comprehension"). Verification research and the construction of mental world models received little attention in early psycholinguistic theory formation which focused rather on parsing principles, the modularity debate, and, gradually, a shift from principle- to lexicon- and, expectation-based approaches to comprehension. Cooper's study and its evidence for rapid listening-looking interaction was acknowledged in publications on the eye-mind hypothesis (Just and Carpenter 1976) but much psycholinguistic theorizing in the late 1980s, and early 1990s seemed unaware of Cooper's findings and their implications. Conversely, the context-focused approaches to comprehension were—except for ECG—underspecified regarding syntactic structure building and semantic interpretation (in an incremental manner).

But each of the context-based strands of research from the 70s and 80—verification, semantic activation, the inferential and script-based construction of mental models of the world, and real-time interaction of language comprehension with visual attention—has

influenced recent psycholinguistic modeling research. Verification mechanisms have been integrated in sentence processing accounts (Knoeferle et al. 2014; Van Herten et al. 2006). Priming research has made its way into psycholinguistic theory in the form of ‘alignment’ between interlocutors (viz. that in dialogue interlocutors align their mental representations Pickering and Garrod 2004). Alignment is postulated to occur at all linguistic levels, up to situation models. The construction of mental models has further influenced the ‘immersed experiencer model’ (Zwaan and Ross 2004) (see Barsalou 1999, for a comprehensive overview of perceptual theories of cognition) and probabilistic neurocomputational models of expectation-based comprehension (Venhuizen et al. 2018). Research on the interaction of language comprehension with visual attention has been continued in speech perception models (Allopenna et al. 1998; Smith et al. 2017), in processing accounts of situated sentence comprehension (Altmann and Kamide 2007; Huettig et al. 2018; Knoeferle and Crocker 2006, 2007), and in computational models of visual attention and situated language comprehension (Crocker et al. 2010; Kukona and Tabor 2011; Mayberry et al. 2009; Roy and Mukherjee 2005).

Three key processing steps and associated variability

From the discussed theoretical approaches, we can identify key processes, and associated mechanisms in comprehension in rich visual context—*sine qua* comprehenders would not be able to *fully* comprehend the meaning of utterances.

- Building structure and assigning an interpretation (by interpretation we mean a mental representation of sentence meaning, informed by linguistic, and world knowledge, and the immediate linguistic and non-linguistic context)
- Grounding the interpretation in a model of the world/visual context via (internal/visual) attention (by ‘model’ we mean mental representations of the immediate non-linguistic context, (see Knoeferle and Crocker 2007; Knoeferle et al. 2014))
- Verifying the structure and interpretation against representations of the world/visual context and revising as necessary

Psycholinguists have argued that each of these three processes—structuring and interpretation, grounding, and verification can vary and apply to different extents regarding depth of interpretation, of grounding and of verification. By contrast, other variability such as that derived from comprehender characteristics and/or world-language relations has only recently been investigated (see “[Variation in expectation-based comprehension and in context effects by comprehender characteristics](#)” and “[Differentiating grounding: do different world-language relations elicit distinct context effects?](#)”).

Depths of structuring and interpretation Regarding structure building and interpretation, Frazier and Fodor (1979) assumed in-depth structure building, as did MacDonald et al. (1994), McRae et al. (1998). By contrast, Ferreira et al. (2002) postulated ‘good-enough’ representations, meaning that comprehenders do not always reconstruct a correct sentence representation. For instance, for *While Bill hunted the deer ran into the woods*, comprehenders tend to initially attach *deer* as the direct object of hunted; at sentence end, however, it should become clear that hunted is used intransitively (p. 373 Christianson et al. 2001). Participants in the experiment by Christianson et al. (2001), however, seemed to retain the initial mis-attachment. When they were asked *Did Bill hunt the deer?* (see Christianson et al. 2001), they responded more often incorrectly that Bill hunted the deer when a sentence initially permitted an object attachment of *deer* to *Bill hunted* than when it did not (e.g., *While Bill hunted the pheasant the deer ran into the woods*). Possibly, late closure effects caused the garden-path representations to linger in working memory with comprehenders constructing perhaps a temporally ordered representation of Bill hunting the deer and then the deer running into the woods (at least this would be a situation model account for the observed findings). For related evidence see research by Kukona and Tabor (2011), Tabor et al. (2004).

Gradients of grounding When considering grounding, Johnson-Laird (1981) postulated that a situation model is only sometimes constructed during comprehension; by contrast, with regard to a theory of concepts (not comprehension), Gallese and Lakoff (2005) postulated “that concepts of a wide variety make direct use of the sensory-motor circuitry of the brain” (see Meteyard et al. 2012; Zwaan 2014, for a review of weakly to strongly embodied approaches).

Since concepts are accessed during comprehension, Gallese's and Lakoff's position would seem to suggest a strongly embodied view of comprehension (e.g., Glenberg and Kaschak 2002, for relevant behavioral evidence). One key notion for grounding is to bring in incrementality, much like in ECG (Bergen and Chang 2005), and in the linguistic focusing hypothesis (Taylor and Zwaan 2008), which predicts that “engagement of the motor system during language comprehension is controlled by the focus of the linguistic message” (p. 143, Zwaan et al. 2010).

Extent of verification Some have argued that verification is ‘part and parcel’ (p. 505, Knoeferle et al. 2011) of language processing. Verification effects in event-related brain potentials to verb-action mismatches emerged prior to any explicit verification, and thus at a point in time when verification would not yet have been necessary; but that evidence came from a study in which participants explicitly assessed sentence veracity via a button press at sentence end. Other studies used comprehension tasks and have reported that comprehenders failed to notice incongruence between a target sentence and their knowledge. Erickson and Mattson (1981), reported the so-called ‘Moses illusion’, which we could characterize as a semantic version of good-enough processing. Participants read out aloud a question such as *How many animals of each kind did Moses take on the Ark?* (p. 540 Erickson and Mattson 1981). They were instructed to answer the question but also informed that questions can contain errors and that they should indicate errors by saying ‘wrong’. In spite of this instruction, the Moses-Ark question was answered incorrectly on 81 percent of the trials (p. 543, participants answered *two* instead of ‘wrong’ and failed to clarify that the person on the ark was named Noah).

Below I discuss further variation in grounding/context effects as a function of comprehender characteristics and world-language relations (“[Variation in expectation-based comprehension and in context effects by comprehender characteristics](#)”). With regard to grounding and context effects, I outline predictions in the form of relative (not absolute) preferences of how representations of visual context affect comprehension (“[Differentiating grounding: do different world-language relations elicit distinct context effects?](#)” and “[Conclusions: predicting \(visual\) context effects](#)”).

Variation in expectation-based comprehension and in context effects by comprehender characteristics

Orthogonal to the depth of interpretation, of grounding, and of verification is the much-debated issue of the extent to which comprehension processes are merely incremental or rather expectation-driven (Huettig 2015; Lau et al. 2006; Pickering and Garrod 2013; Pickering and Gambi 2018). A postulate of strong expectation-based comprehension can be derived from surprisal theory concerning syntactic processes (Levy 2008). For instance, if a comprehender encounters a sentential subject, the expectation of seeing that constituent again in the same clause decreases following the logic that multiple same-type constituents rarely co-occur in a clause (see p. 1146 Levy 2008). A strong view of expectations is also compatible with evidence suggesting highly specific expectations of lexical elements following a constraining sentence context and determiner (*a* vs. *an* as in *The boy went to the park and flew a/an kite/airplane* (DeLong et al. 2005). Readers exhibited larger mean amplitude negativities when they encountered *an* than *a*, a finding which was attributed to comprehenders noticing an incongruence between their expectation of a plausible noun—starting with a consonant—such as *kite* and the article *an*. DeLong et al. (2005) concluded that “predictions can be for specific phonological forms—words beginning with either vowels or consonants. In this sense, we propose that prediction can be highly specific, at least under some circumstances” (p. 1119f.). The N400 is, however, modulated by many further factors, among them plausibility and semantic relatedness (Nieuwland 2019); there is further ongoing debate as to how specific (word form or meaning) (Nieuwland et al. 2018), and how robust these expectations effects are (DeLong et al. 2017; Ito et al. 2017).

Federmeier (2007) explicitly points out variability in expectations and suggests that “... the brain uses context to predict features of likely upcoming items. However, although prediction seems important for comprehension, it also appears susceptible to age-related deterioration and can be associated with processing costs” (p. 491). That specific expectations are not always realized is underscored by recent replication failures of DeLong and colleagues’ finding (DeLong et al. 2017; Nieuwland et al. 2018; Ito et al.

2017), and by the insight that anticipatory processes (how quickly comprehenders visually anticipate an object before its mention given constraining linguistic context) varies by characteristics of the comprehender. Among these are his/her literacy (Huettig et al. 2018; Mishra et al. 2012), language production skills (Mani and Huettig 2012), working memory and processing speed (Huettig and Janse 2015), reading skills (Huettig and Brouwer 2015), and native (vs. non-native) language command (Ito et al. 2018). These findings extend to age-related variation in the time course of context effects (Münster 2016; Münster and Knoeferle 2018, for a related account). All of these studies share that they manipulate comprehender characteristics by comparing comprehension in different participant groups. Mishra et al. (2012), for instance, presented spoken sentences that did (vs. didn't) contain highly constraining words (e.g., the adjective *uncha/i*, 'high', and a particle *wala/i*) to high and low literates. At issue was whether the semantic and syntactic (gender) constraints provided by *uncha/i wala/i* would prompt comprehenders to anticipate the target object (a door) more than distractor objects incompatible with the semantic and syntactic constraints. Only high, but not low literates, visually anticipated the target door before its mention, a finding that the authors attribute to differences in the mental representations and / or processes in low (vs. high) comprehenders. In Mani and Huettig (2012), only children with a large (vs. smaller) vocabulary at age two, engaged in such anticipatory gaze behavior. Ito et al. (2018) observed that speaker competence appeared to modulate the visual anticipation of phonological competitor objects. Only English native speakers but not second-language speakers (first language: Japanese) anticipated the referent of an English-language phonological competitor (clown) when the context constrained meaning towards a highly likely target object (*cloud* after *The tourists expected rain when the sun went behind the...* but see Chambers and Cooke 2009, for failure to find related effects of language proficiency).

Differentiating grounding: do different world-language relations elicit distinct context effects?

The focus in context-based approaches to comprehension (see “[Context-focused approaches to comprehension and associated representational assumptions](#)”) has been on modeling comprehension

with an average comprehender in mind, and without further differentiating the notion of contextual effects (unlike the distinction of syntactic from semantic processes and investigations of their interplay, see Hagoort 2003). Situation models contained different aspects of context (see also Jackendoff 2002), but it is important to also consider distinct world-language relations (e.g., referential vs. associative, referential vs. role relations), as well as speakers and their gaze and facial mimics as part of the visual context (Knoeferle and Guerra 2012).

That visual context effects are generic is not a claim that has been made explicitly in the literature. However, it appears to be assumed tacitly via labeling context effects generically as ‘visual context effects’ or ‘visual information’ (e.g., Tanenhaus et al. 1995; Huettig et al. 2018; Knoeferle and Crocker 2006; Sedivy et al. 1999). It has been shown that virtually any kind of information in the visual context rapidly modulates language comprehension. Among these are object properties such as size (Sedivy et al. 1999), shape Dahan and Tanenhaus (2005), and color (Huettig et al. 2018). Rapid effects on visual attention and language comprehension have also been reported for action depictions (Knoeferle et al. 2005, 2011), action affordances (Chambers et al. 2004), visual saliency (Coco and Keller 2015), speaker gaze (Hanna and Brennan 2007; Knoeferle and Kreysa 2012), and speaker facial emotions (Carminati and Knoeferle 2013). Parsimony dictates to assume a single mechanism underlying all of these effects. But recent evidence suggests we need a more fine-grained (and perhaps more principled) accounts of context effects to capture the observed variability. Some evidence for distinction between context effects has been contributed, for instance, by visual-world research reported in Huettig and McQueen (2007): these studies observed a temporal cascade in the activation of phonological, shape, and semantic competitors during spoken comprehension that varied moreover as a function of the preview time.

Robust referential and action effects: a processing preference? This variability notwithstanding, action effects, for example, have been robustly observed in numerous studies, using both eye-tracking and event-related brain potentials as measures; the effects emerged as soon as the verb mediated an action (Knoeferle et al. 2005; Knoeferle and Crocker 2006, 2007; Knoeferle et al. 2008, 2011, 2014). For

instance, when comprehenders encountered a verb (e.g., ‘paints’) in the structurally ambiguous spoken sentence ‘The princess (amb.: subj/obj?) paints ...’; example translated from German), they rapidly related it to an action event depicting a princess as being painted by a fencer, eliciting resolution of the temporary ambiguity towards the non-canonical structure with the princess as the object and patient of the event. Action event depiction influenced both syntactic disambiguation (Knoeferle et al. 2005, 2008), and semantic processes implicated in relating an action to the verb (Knoeferle et al. 2011). Processes of establishing reference, also seem to be rapid and robust, as evidenced in the findings by Cooper (1974), Tanenhaus et al. (1995) and Sedivy et al. (1999).

Evidence for preferences in cue processing emerged when comparing the effects of referentially-mediated action events with effects that were mediated less directly (stereotypical associations between a verb and an agent). When a verb like *bandages* mediated a bandaging action and its agent (a chef) in a clipart scene *Den Touristen bandagiert gleich der ...*, ‘The tourist (obj) bandages soon the ... (subj)’, comprehenders anticipated the referentially-mediated action and the chef more before its mention than a competing stereotypical agent, a medico (Knoeferle and Crocker (2006).

These findings support the view that not all language-world relations are processed similarly (see also Cooper 1974). Different aspects of context can be distinguished based on their relationship to language (e.g., referential vs. non-referential world-language relations). Cooper (1974) reported that participants inspected objects that were mediated by language via referential and superordinate-token relations (e.g., *forrest—tree*) more often than objects that were mediated otherwise by language (e.g., in *The queen was in agony*, *agony* was indirectly related to a picture of a queen, p. 88). Some support for a referential preference also comes from a comparison of looks to target objects [e.g., a piano when hearing *piano* in *Eventually, the man agreed hesitantly but then he looked at the piano and appreciated that it was beautiful* compared with an also-present trumpet (p. B25 Huettig and Altmann 2005)], suggesting referential relations trump non-referential ones. In another study (Kukona et al. 2014), participants listened to *The boy will eat the white ...* while inspecting a display depicting a white cake, a brown cake, a white

car, and a brown car. In that setting, participants inspected the white cake more than the brown cake, suggesting verb constraint and adjective constraint are rapidly integrated. However, they also inspected the white *car* more than the brown car. Thus local referential constraints influenced attention and comprehension (a white car was considered as a potential referent) even when verb-based anticipatory constraints should have excluded cars as referents.

Distinguishing referential from non-referential relations, and viewing the former as more central enables us to derive predictions about how rapidly and strongly a particular aspect of context should modulate comprehension. That reasoning receives support from arguments of core (vs. peripheral) semantic relations. For instance, the word *zebra* identifies the animal zebra and conjures up memories (even if imperfect) of what zebras look like and as such the relation between *zebra* and the animal zebra are core (see Rosch 1973, for related notions on category prototypicality). By contrast, *feeding*, or *the zookeeper* might conjure up images of a zebra (especially if a zebra is nearby) but in contrast to referential relations, the relation would be more peripheral. Similarly, *smell* would be associated with a nose but the relation is less core than between the word *nose* and its referent (see Duñabeitia et al. 2008, for related evidence). Referential cueing of information in visual context, on this account, would on average elicit stronger effects than non-referential mediation of visual context. This preference is likely not absolute, but we can use it as a test case for hypotheses about what governs rapid context effects during language comprehension.

Actions versus speaker gaze in comprehension and recall Further evidence that distinct language-world relations elicit distinct context effects come from eye-tracking in visual contexts. For instance, Kreysa et al. (2018) compared verb-mediated depicted action effects with speaker gaze effects in a crossed, two-by-two design (actions cueing a target were present vs. absent; the speaker gazed at the target object or was obscured). Such a direct experimental comparison permits characterizing the relative contribution of distinct language-world relations to situated language comprehension.

Kreysa et al. (2018) monitored participants’ eye movements to characters on a computer display while the participants listened to transitive agent-action-patient (subject-verb-object) sentences. In two eye-

tracking experiments, the authors manipulated whether speaker gaze, a depicted action, neither, or both of these were visible. Across the experiments either both cues were deictic world-language relations (Experiment 1) or only speaker gaze (Experiment 2) was. This design permitted teasing apart the contribution of deictic relations (both actions and gaze were used in a deictic manner in Experiment 1) from that of cue type (action vs. speaker gaze). The gaze record revealed that deictic cues (speaker gaze and a single deictic action) modulated comprehension rapidly, and significantly earlier than a non-deictic action depiction. Not only did these two information types (actions, speaker gaze) differ in the effects on real-time language comprehension (reflected in visual attention) but also in how they affected ensuing recall of sentence content (the object was recalled better when an action depiction had been present vs. absent). Together these findings suggest that information type and its relation to language characterize real-time situated language comprehension and recall of sentence content.

Interesting insights into potentially distinct effects of cues such as a speaker's gaze and an arrow (that can point to something in a deictic fashion) come also from research by Staudte et al. (2014). They compared the effects of a speaker's gaze and of an arrow in referential processing and observed that when they are matched in visual precision, their effects were comparable (Staudte et al. 2014). This can be taken to support the view that similar relation of (cues in) the world to conveyed language results in similar effects. Showing that when precision (Staudte et al. 2014) or language-world relation (Kreysa et al. 2018) are matched, cue effects on real-time comprehension are similar, is noteworthy. However, one might argue that cues in context differ at least sometimes in their appearance, timing, and relation to language and that we must derive a principled account of what language-world relation elicits what effect on language comprehension (its time course and representations). To the extent that world-language relations govern context effects, the prediction from these findings is that verb-mediated action effects should influence comprehension and recall of the interpretation content more than cues that are not referentially tied into the interpretation [e.g., speaker gaze, relations between *the queen was in agnony* and a depiction of the queen, emotional facial expressions cued by adjectives such as *happy*, distance between objects related to semantic

similarity (Guerra and Knoeferle 2014)]. Kukona et al. (2011) add evidence to the argument of distinct language-world relations modulating context effects: As participants in their studies listened to sentences such as *Toby arrested the thief.*, they inspected both a picture of a thief and of a policeman to a similar extent and more than distractor objects during *arrested the*. Looks to the thief only exceeded looks to the policeman as *the thief* was processed. The latter result suggests that sometimes lexical-semantic associations may rapidly guide attention to objects. One interpretation of this gaze pattern is that both policeman and thief were active as part of an event, eliciting inspection; alternatively, as the authors suggested, local constraints imposed by the verb may modulate compositional semantic interpretation [if eye movements follow the unfolding syntactic structure and semantic–thematic interpretation, then strictly speaking, a policeman should not have received a noticeable amount of attention (he is unlikely the patient of an arresting action)].

When events contribute to referential vs. thematic role relation processes Context effects on language comprehension differ not only by world-language relation but also by what sub-processes in comprehension they inform. Different types of picture–sentence mismatches, for instance, implicated different brain responses as indexed by ERP effects (Knoeferle et al. 2014). Participants read sentences such as *The gymnast punches the journalist*, after they had seen a clipart depiction that either fully matched the sentence, mismatched in the action, the role-relations, or both of these. If context effects implicated a single mechanism only, then we should see comparable mismatch effects in comprehenders' brain responses. However, this was not the case: When role relations mismatched, effects emerged at *journalist* as anterior negativities, preceding centro-parietal N400 action mismatch effects at *punches*. Post-verbally, these mismatch manipulations also yielded different effects, and correlated differently with a participant's mean accuracy in the verification task, verbal working memory and visual–spatial scores, and differed in their interactions with stimulus onset asynchrony (see also Wassenaar and Hagoort 2007).

These results were interpreted as implicating more than a single mismatch mechanism during comprehension. When comparing these results to those by Knoeferle et al. (2008), it is striking that in the latter

study, linguistic cues (case marking) and event depictions seemed to affect comprehension similarly (as reflected in similar ERP effects), perhaps because these different information types contributed to one and the same—structural disambiguation—process. By contrast, in Knoeferle et al. (2014), the action and role relation depictions contributed to distinct comprehension processes (figuring out who-does-what-to-whom and grounding the action). From this we can derive a further testable prediction, viz, that when distinct or similar cues contribute to the same comprehension process, they should affect comprehension in a similar manner; by contrast, if they inform distinct comprehension processes, their effects will show up as distinct.

Conclusions: predicting (visual) context effects

The present review identified three predictions concerning visual context effects in an attempt to derive a more principled account of context effects:

- P1 Comprehender characteristics e.g., age, literacy, language skills modulate context effects in language comprehension (“[Variation in expectation-based comprehension and in context effects by comprehender characteristics](#)”)
- P2 Referential cueing of information in visual context elicits—on average and by comparison—more rapid and stronger effects than non-referential mediation of visual context (“[Differentiating grounding: do different world-language relations elicit distinct context effects?](#)”); world-language relations thus predict context effects.
- P3 When distinct or similar cues contribute to the same comprehension process, they should affect comprehension in a similar manner; by contrast, if they inform distinct comprehension processes, their effects will show up as distinct (“[Differentiating grounding: do different world-language relations elicit distinct context effects?](#)”).

While language comprehension in visual context can clearly vary as a function of comprehender characteristics, the reviewed accounts do not explicitly integrate speaker and comprehender characteristics as a

modulatory factor in language comprehension. Moreover, they are underspecified regarding how grounding and context effects influence comprehension. Much psycho- and neurolinguistic research has, however, assessed to what extent language comprehension depends on the perceived context, the body, and long-term linguistic knowledge of the comprehender (e.g., Barsalou 2008; Hasson et al. 2018; Holcomb et al. 1992; Kotz 2009; Osterhout et al. 2008; Zwaan and Ross 2004; Zwaan 2016) among others. Extant approaches have constrained predictions of context effects in language processing among others via

1. a hierarchy of physical world, an agent’s body, and context in the ‘TEST’ framework (Myachykov et al. 2014)
2. levels of embodiment from demonstration to abstraction (Zwaan 2014, p. 229)
3. the presence of working memory representations, capacity limits, and cognitive control (Huettig et al. 2011), and
4. a referential mechanism and decay in working memory (Knoeferle and Crocker 2006, 2007), verification mechanisms (Knoeferle et al. 2014), and speaker/listener characteristics (Münster and Knoeferle 2018) in the (social) coordinated interplay account

Situation-based and TEST-based prediction of context effects

Zwaan (2014) differentiates five forms of language comprehension depending on overlap between the context and the communicative situation. In a first high-overlap situation, agents, objects, and actions are present in the situation (‘demonstrations’, e.g., cooking); in a second situation, objects are not immediately present such as when giving instructions to someone to fetch an object. Even less embedded are projections (mapping a past state of the context onto the present context, e.g., a builder explaining the changes to a house following renovation, p. 232), followed by displacements (a context unrelated to the present context), and abstractions (e.g., scientific articles). This situation-based approach would predict strong context effects in language comprehension for referential and instruction situations, with diminishing context effects when the referential context and the communicative situation are less overlapping. The

approach would not be able to make finer-grained predictions such as whether within one such situation, referential relations influence language comprehension more than other world-language relations.

Myachykov et al. (2014) present a framework of embodied representations and distinguish between invariant aspects of the world such as gravity (p. 446) from somewhat stable embodied aspects (e.g. of an agent's state such as frontal vision, p. 446) to less stable aspects such as an agent's interpretation of a specific environment given her goals. Myachykov et al. (2014) view these aspects as hierarchical, and differentially susceptible to learning. The framework can make predictions about context effects when comparing the relative influence of physical aspects of the world compared with aspects of the immediate context or an agent's body. To accommodate distinct language-world relations, we would, in addition, need the context part of the framework to be specified further.

Working memory as a hub in language-mediated visual attention

Huettig et al. (2011) focus on accommodating listening-looking interactions, and posit that “first the visual display is processed up to a high level, including the creation of conceptual and linguistic representations. At this high level, these representations subsequently match up with those activated by the linguistic input, activation that then feeds back to the linked location” (p. 142 Huettig et al. 2011). Huettig et al. (2011) posit that working memory is central in accommodating language-mediated attention to visual (object) contexts.

From that postulate, they derive a number of predictions such as that working memory representations should suffice to guide attention to objects during comprehension and this may happen automatically. A further prediction was that active working memory representations are needed for attentional guidance, and that long-term memory can also guide attention. The discussed evidence supports the view that representations in working memory are necessary to guide attention but that long-term memory may also influence attention to things. For situated language comprehension, the authors argue that with enough time, comprehenders link semantic and phonological representations to an object and its location (see related

accounts of visual search and language-mediated situation model construction Spivey et al. 2001; Spivey and Geng 2001). This approach might be able to accommodate distinct context effects—provided that they fall out of the activation of representations in working memory. To the extent that comprehender characteristics modulate context effects via working or long-term memory, the account might be able to predict comprehender-specific modulation of context effects (but may require further refinement).

Comprehender characteristics and processing preferences

In relation to these proposals, the present approach focuses on differentiating context effects in language comprehension via a characterization of more or less core world-language relations; on the effects of comprehenders characteristics (see Münster and Knoeferle 2018, for more detail); and on a more principled account of (mostly visual) context effects, relying on the predictions (P1)–(P3). I have proposed to foreground distinct world-language relations and comprehender characteristics to predict (variability in) context effects. Comprehender characteristics have been shown to cause variability in context effects. Comprehender characteristics such as literacy would on that account set the probability of expecting a specific word such as ‘door’ to a low value for low literates, and to a high value for high literates, effectively capturing the absence of expectations and visual anticipation in low literates (see p. 7, Münster and Knoeferle 2018).

Concerning world-language relations, not every aspect of (visual) context influences language comprehension in the same manner. Indeed, how language mediates context information (e.g., via reference or other links), appears to predict (variability in) context effects. Processing preferences come into play in that some context effects (e.g., of verb-mediated actions and referential relations) seem particularly robust and pervasive—in reading and listening, across real-world, video-taped, and depicted environments, and across children, young, and older adults, suggesting generalizability beyond individual variation.

An open issue is how to rank these two predictors (comprehender characteristics and referential preferences)—would they be on equal footing, or rather hierarchical in their influence on real-time

comprehension? The account by Münster and Knoeferle (2018) predicts that comprehender characteristics would modulate context effects [see Figure 1 in Münster and Knoeferle (2018), where comprehender characteristics can modulate the content of working memory and associated probabilistic expectations].

Acknowledgements This research has received support from funding by the German Research Council, FoTeRo (SPP 1727 Xprag). The corresponding author states that there is no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Alloppenna, P., Magnuson, J., & Tanenhaus, M. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory & Language*, 38, 419–439.
- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73, 247–264.
- Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, 57, 502–518.
- Amsel, B. D., Urbach, T. P., & Kutas, M. (2014). Empirically grounded cognition: The case of color. *NeuroImage*, 99, 149–157.
- Arbib, M. A., & Lee, J. (2008). Describing visual scenes: Towards a neurolinguistics based on construction grammar. *Brain Research*, 1225, 146–162.
- Barsalou, L. W. (1999). Perceptual and symbol systems. *Behavioural and Brain Sciences*, 22, 577–609.
- Barsalou, L. W. (2008). Cognitive and neural contributions to understanding the conceptual system. *Current Directions in Psychological Science*, 17, 91–95.
- Bates, E., Friederici, A., & Wulfeck, B. (1987). Comprehension in aphasia: A cross-linguistic study. *Brain and Language*, 32, 19–67.
- Bergen, B., & Chang, N. (2005). Embodied construction grammar in simulation-based language understanding. In J.-O. Östman & M. Fried (Eds.), *Construction Grammars: Cognitive grounding and theoretical extensions* (pp. 147–190). Amsterdam: John Benjamins.
- Bock, J. (1986). Syntactic persistence in language processing. *Cognitive Psychology*, 18, 355–387.
- Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review*, 113, 787–821.
- Bryant, J. (2008). Best-fit constructional analysis. Unpublished doctoral dissertation, Berkeley, University of California Berkeley.
- Carminati, M. N., & Knoeferle, P. (2013). Effects of speaker emotional facial expression and listener age on incremental sentence processing. *PLoS One*, 8, e72559. <https://doi.org/10.1371/journal.pone.0072559>.
- Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review*, 82, 45–73.
- Chambers, C. G., & Cooke, H. (2009). Lexical competition during second-language listening: Sentence context, but not proficiency, constrains interference from the native lexicon. *JEP: LMC*, 35, 1029–1040.
- Chambers, C. G., Tanenhaus, M. K., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains during real time language comprehension. *Journal of Memory and Language*, 47, 30–49.
- 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, 687–696.
- Christianson, K., Hollingworth, A., Halliwell, J., & Ferreira, F. (2001). Thematic roles assigned along the garden path linker. *Cognitive Psychology*, 42, 368–407.
- Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. *Perspectives on Socially Shared Cognition*, 13(1991), 127–149.
- Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology*, 3, 472–517.
- Clifton, C., Frazier, L., & Connine, C. (1984). Lexical expectations in sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, 23, 696–708.
- Coco, M., & Keller, F. (2015). The interaction of visual and linguistic saliency during syntactic ambiguity resolution. *Quarterly Journal of Experimental Psychology*, 68, 46–74.
- Coltheart, M. (1999). Modularity and cognition. *Trends in Cognitive Science*, 3, 115–120.
- Connell, L. (2007). Representing object color in language comprehension. *Cognition*, 102, 476–485.
- Cooper, R. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, 6, 84–107.
- Crocker, M. W., Knoeferle, P., & Mayberry, M. (2010). Situated sentence comprehension: The coordinated interplay account and a neurobehavioral model. *Brain and Language*, 112, 189–201.
- Dahan, D., & Tanenhaus, M. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review*, 12, 453–459.
- D'Arcy, R., & Connolly, J. (1999). An event-related brain potential study of receptive speech comprehension using a modified token test. *Neuropsychologia*, 37, 1477–1489.

- Delogu, F., Crocker, M. W., & Brouwer, H. (2017). Teasing apart coercion and surprisal: Evidence from eye-movements and ERPs. *Cognition*, 161, 46–59.
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8, 1117–1121.
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2017). Is there a replication crisis? Perhaps. Is this an example? No: A commentary on Ito, Martin, and Nieuwland (2016). *Language, Cognition, and Neuroscience*, 96, 367–400. <https://doi.org/10.1016/j.neubiorev.2018.11.019>.
- Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109, 193–210.
- Duñabeitia, A., Avilés, A., Afonso, O., Scheepers, C., & Carreiras, M. (2008). Qualitative differences in the representation of abstract versus concrete words: Evidence from the visual-world paradigm. *Cognition*, 110, 284–292.
- Erickson, T., & Mattson, M. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, 20, 540–551.
- Federmeier, K. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44, 491–505.
- Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11, 11–15.
- Fitneva, S., & Spivey, M. J. (2005). Context and language use: The effect of authorship. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions* (pp. 317–328). London: Bradford Book.
- Fodor, J. (1983). *Modularity of mind*. Cambridge, MA: MIT Press.
- Frank, S. (2013). Uncertainty reduction as a measure of cognitive processing load in sentence comprehension. *Topics in Cognitive Science*, 5, 475–494.
- Frank, S., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, 140, 1–11.
- Frazier, L., & Fodor, J. D. (1979). The sausage machine: A new two-stage parsing model. *Cognition*, 6, 291–325.
- Friederici, A. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6, 78–84.
- Gallese, V., & Lakoff, G. (2005). The Brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, 22, 455–479.
- Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for “common sense”: An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience*, 8, 89–106.
- Garnsey, S. M., Pearlmutter, N. J., Myers, E. M., & 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.
- Glenberg, A., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin and Review*, 9, 558–565.
- Goldberg, A. (1995). *Construction grammar: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Goldberg, A. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Goolkasian, P. (1996). Picture-word differences in a sentence verification task. *Memory & Cognition*, 24, 584–594.
- Gough, P. B. (1965). Grammatical transformations and speed of understanding. *Journal of Verbal Learning and Verbal Behavior*, 5, 107–111.
- Grossman, M., Carvell, S., Stern, M. B., Gollomp, S., & Hurtig, H. I. (1992). Sentence comprehension in Parkinson's disease: The role of attention and memory. *Brain and Language*, 42, 347–384.
- Grossman, M., Carvell, S., Stern, M. B., Reivich, M., Morrison, D., Alavi, A., et al. (1993). Cognitive and physiological substrates of impaired sentence processing in Parkinson's disease. *Journal of Cognitive Neuroscience*, 5, 480–498.
- Guerra, E., & Knoeferle, P. (2014). Spatial distance effects on incremental semantic interpretation of abstract sentences: Evidence from eye tracking. *Cognition*, 133, 535–552.
- Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience*, 15, 883–899.
- Hale, J. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research*, 32(1), 101–122.
- Hanna, J. E., & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory & Language*, 57, 596–615.
- Hasson, U., Egidi, G., Marelli, M., & Willems, R. (2018). Grounding the neurobiology of language in first principles. *Cognition*, 180, 135–157.
- Holcomb, P. J., Coffey, S. A., & Neville, H. J. (1992). Visual and auditory sentence processing: A developmental analysis using event-related brain potentials. *Developmental Neuropsychology*, 8, 203–241.
- Huetting, F. (2015). Four central questions about prediction in language processing. *Brain Research*, 1626, 118–135.
- Huetting, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, 96, B23–B32.
- Huetting, F., & Brouwer, S. (2015). Delayed anticipatory spoken language processing in adults with dyslexia—Evidence from eye-tracking. *Dyslexia*, 21, 97–122.
- Huetting, F., & Janse, E. (2015). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition, and Neuroscience*, 31, 80–93.
- Huetting, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, 57, 460–482.
- Huetting, F., Olivers, C., & Hartsuiker, R. (2011). Looking, language, and memory: Bridging research from the visual world and visual search paradigms. *Acta Psychologica*, 137, 138–150.

- Huetting, F., Singh, N., & Mishra, R. K. (2018). Language-mediated visual orienting behavior in low and high literates. *Frontiers in Psychology*, 2, 285. <https://doi.org/10.3389/fpsyg.2011.00285>.
- Ito, A., Martin, A., & Nieuwland, M. (2017). Why the a/an prediction effect may be hard to replicate: a rebuttal to Delong, Urbach, and Kutas (2017). *Language, Cognition, and Neuroscience*, 32, 974–983.
- Ito, A., Pickering, M., & 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, 1–11.
- Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.
- Johnson-Laird, P. N. (1981). Comprehension as the construction of mental models. *Philosophical Transactions of the Royal Society London: B*, 295, 353–374.
- Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, 8, 441–480.
- Kimball, J. (1973). Seven principles of surface structure parsing in natural languages. *Cognition*, 2, 15–47.
- King, J. W., & Kutas, M. (1995). Who did what and when? Using word- and clause-level erps to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, 7, 376–395.
- Knoeferle, P., & Crocker, M. W. (2005). Incremental effects of mismatch during picture-sentence integration: Evidence from eye-tracking. In *Proceedings of the 27th Annual Meeting of the Cognitive Science Conference* (pp. 1166–1171). Mahwah, NJ: Erlbaum.
- Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science*, 30, 481–529.
- Knoeferle, P., & Crocker, M. W. (2007). The influence of recent scene events on spoken comprehension: Evidence from eye-movements. *Journal of Memory and Language*, 75, 519–543.
- 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, 95–127.
- Knoeferle, P., & Guerra, E. (2012). What's non-linguistic visual context? A view from language comprehension. In R. Finkbeiner, J. Meibauer, & P. Schumacher (Eds.), *What is a context? Linguistic approaches and challenges*. (pp. 129–150). Amsterdam: John Benjamins.
- Knoeferle, P., Habets, B., Crocker, M. W., & Münte, T. F. (2008). Visual scenes trigger immediate syntactic reanalysis: Evidence from ERPs during situated spoken comprehension. *Cerebral Cortex*, 18, 789–795.
- Knoeferle, P., & Kreysa, H. (2012). Can speaker gaze modulate syntactic structuring and thematic role assignment during spoken sentence comprehension? *Frontiers in Psychology*, 3, 538. <https://doi.org/10.3389/fpsyg.2012.00538>.
- Knoeferle, P., Urbach, T. P., & Kutas, M. (2011). Comprehending how visual context influences incremental sentence comprehension: Insights from ERPs and picture-sentence verification. *Psychophysiology*, 48, 495–506.
- 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.
- Kotz, S. A. (2009). A critical review of ERP and fMRI evidence on L2 syntactic processing. *Brain & Language*, 109, 68–74.
- Kreysa, H., Nunnemann, E. M., & 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, 220–229.
- Kukona, A., Cho, P. W., Magnuson, J., & 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, 326–347.
- Kukona, A., Fang, S.-Y., Aicher, K., Chen, H., & Magnuson, J. (2011). The time course of anticipatory constraint integration. *Cognition*, 119, 23–42.
- Kukona, A., & Tabor, W. (2011). Impulse processing: A dynamical systems model of incremental eye movements in the visual world paradigm. *Cognitive Science*, 35, 1009–1051.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, 62, 621–647.
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307, 161–163.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.
- Langacker, R. W. (1987). *Foundations of cognitive grammar* (Vol. 2.). Descriptive application Stanford, CA: Stanford University Press.
- Lau, E., Stroud, C., Plesch, S., & Phillips, C. (2006). The role of structural prediction in rapid syntactic analysis. *Brain and Language*, 98, 74–88.
- Levy, R. (2008). Expectations-based syntactic comprehension. *Cognition*, 106, 1126–1177.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676–703.
- MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behaviour*, 23, 127–150.
- Mani, N., & Huetting, F. (2012). Prediction during language processing is a piece of cake—But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, 38, 833–847.
- Marslen-Wilson, W. D. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, 244, 522–523.
- Mayberry, M., Crocker, M. W., & Knoeferle, P. (2009). Learning to attend: A connectionist model of situated language comprehension. *Cognitive Science*, 33, 449–496.
- McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38, 283–312.
- Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, 48, 788–804.

- Mishra, R. K., Singh, N., Pandey, A., & Huettig, F. (2012). Spoken language-mediated anticipatory eye-movements are modulated by reading ability—Evidence from indian low and high literates. *Journal of Eye Movement Research*, 5, 1–10.
- Mitchell, D. C., Cuertos, F., Corley, M., & Brysbaert, M. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research*, 24, 469–488.
- Münster, K. (2016). Effects of emotional facial expressions and depicted actions on situated language processing across the lifespan. Unpublished doctoral dissertation, Bielefeld University, Bielefeld University, Bielefeld, Germany. <https://pub.uni-bielefeld.de/record/2906648>. Accessed 19 Dec 2017.
- Münster, K., & Knoeferle, P. (2018). Extending situated language comprehension (accounts) with speaker and comprehender characteristics: Toward socially situated interpretation. *Frontiers in Psychology*, 8, 2267. <https://doi.org/10.3389/fpsyg.2017.02267>.
- Myachykov, A., Scheepers, C., Fischer, M., & Kessler, K. (2014). Test: A tropic, embodied, and situated theory of cognition. *Topics in Cognitive Science*, 6, 442–460.
- Nichol, J. L., & Pickering, M. J. (1993). Processing syntactically ambiguous sentences: Evidence from semantic priming. *Journal of Psycholinguistic Research*, 22, 207–237.
- Nieuwland, M. (2019). Do ‘early’ brain responses reveal word form prediction during language comprehension? A critical review. *Neuroscience and Biobehavioral Reviews*, 96, 367–400. <https://doi.org/10.1016/j.neubiorev.2018.11.019>.
- Nieuwland, M., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., et al. (2018). Limits on prediction in language comprehension: A multi-lab failure to replicate evidence for probabilistic pre-activation of phonology. *eLife*, 7, 33468. <https://doi.org/10.7554/eLife.33468>.
- Osterhout, L., Poliakov, A., Inoue, K., McLaughlin, J., Valentine, G., Pitkanen, I., et al. (2008). Second language learning and changes in the brain. *Journal of Neurolinguistics*, 21, 509–521.
- Pickering, M., & Garrod, S. (2004). Towards a mechanistic psychology of dialogue. *Brain and Behavioral Sciences*, 27, 169–226.
- Pickering, M., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Brain and Behavioral Sciences*, 36, 329–347.
- Pickering, M., & Traxler, M. (1995). Plausibility and the recovery from garden-paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 940–961.
- Pickering, M. J., & Branigan, H. (1999). Syntactic priming in language production. *Trends in Cognitive Sciences*, 3, 136–141.
- Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, 144, 1002–1044.
- Potter, M. C., & Faulconier, B. A. (1975). Time to understand pictures and words. *Nature*, 253, 437–438.
- Rosch, E. (1973). Natural categories. *Cognitive Psychology*, 4, 328–350.
- Roy, D., & Mukherjee, N. (2005). Towards situated speech understanding: Visual context priming of language models. *Computer Speech and Language*, 19, 227–248.
- Schank, R., & Abelson, R. (1975). Scripts, plans, and knowledge. In *Proceedings of the 4th international joint conference on artificial intelligence* (pp. 151–157). Tbilisi, USSR: Morgan Kaufmann Publishers.
- Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, 71, 109–148.
- Singer, M. (2006). Verification of text ideas during reading. *Journal of Memory and Language*, 54, 574–591.
- Smith, A. C., Monaghan, P., & Huettig, F. (2017). Complex word recognition behaviour emerges from the richness of the word learning environment. In A. C. S. K. Twomey, G. Westermann, & P. Monaghan (Eds.), *Proceedings of the 14th neural computation and psychology workshop* (pp. 99–114). Singapore: World Scientific.
- Sperber, R. D., McCauley, C., Ragain, R. D., & Weil, C. M. (1979). On computational and behavioral evidence regarding hebbian transcortical cell assemblies. *Memory & Cognition*, 7, 339–345.
- Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research*, 65, 235–241.
- 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, 447–481.
- Spivey, M. J., Tyler, M. J., Eberhard, K. M., & Tanenhaus, M. K. (2001). Linguistically mediated visual search. *Psychological Science*, 12, 282–286.
- Staudte, M., Crocker, M. W., Heloir, A., & Kipp, M. (2014). The influence of speaker gaze on listener comprehension: Contrasting visual versus intentional accounts. *Cognition*, 133, 317–328.
- Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, 50, 355–370.
- 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, 310–317.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 632–634.
- Taylor, L. J., & Zwaan, R. A. (2008). Motor resonance and linguistic focus. *Quarterly Journal of Experimental Psychology*, 61, 896–904.
- Trijp, R. V., Steels, L., Beuls, K., & Wellens, P. (2012). Fluid construction grammar: The new kid on the block. In *Proceedings of the 13th conference of the European chapter of the association for computational linguistics* (pp. 63–68).
- Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language*, 33, 285–318.
- Underwood, G., Jebbett, L., & Roberts, K. (2004). Inspecting pictures for information to verify a sentence: Eye movements in general encoding and in focused search. *The*

- Quarterly Journal of Experimental Psychology*, 56, 165–182.
- Van Herten, M., Chwilla, D. J., & Kolk, H. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitive Neuroscience*, 18, 1–17.
- Venhuizen, N., Crocker, M. W., & Brouwer, H. (2018). Expectation-based comprehension: Modeling the interaction of world knowledge and linguistic experience. *Discourse Processes*, 1, 1–27. <https://doi.org/10.1080/0163853X.2018.1448677>.
- Wassenaar, M., & Hagoort, P. (2007). Thematic role assignment in patients with Broca's aphasia: Sentence-picture matching electrified. *Neuropsychologia*, 45, 716–740.
- Zwaan, R. (2014). Embodiment and language comprehension: Reframing the discussion. *Trends in Cognitive Sciences*, 18, 229–234.
- Zwaan, R. (2016). Situation models, mental simulations, and abstract concepts in discourse comprehension. *Psychonomic Bulletin & Review*, 23, 1028–1034.
- Zwaan, R., Taylor, L., & de Boer, M. (2010). Motor resonance as a function of narrative time: Further tests of the linguistic focus hypothesis. *Brain and Language*, 112, 143–149.
- Zwaan, R. A., & Ross, B. (2004). The immersed experiencer: Towards an embodied theory of language comprehension. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp. 35–62). New York: Academic Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.