Computational Psycholinguistics

Matthew W. Crocker

Harm Brouwer

Department of Language Science and Technology, Saarland University

Campus, Building C7.1

66123, Saarbruecken, Germany

To appear in: Ron Sun (ed.), *Cambridge Handbook of Computational Cognitive Sciences*.

CUP: Cambridge, UK.

Computational Psycholinguistics

1. Introduction

How is it that people map between a linguistic signal and a mental representation of the meaning that signal encodes? While this mapping can be viewed from both the language production (see Dell & Cholin, 2012) and comprehension perspectives, the focus of this chapter will be on the latter. Even within comprehension, there are numerous stages involved in this process – recovering a phonological or orthographic representation of words, determining their relevant morphological and syntactic properties, retrieving their meaning from long-term semantic memory, and then combining words to recover the intended message of the entire utterance. The complexity, ambiguity and context-dependent nature of language, combined with the dynamical nature of the processes that support comprehension in real time, highlights the need for computational theories – which can be instantiated as implemented computational models – of how people retrieve the words of an utterance as they are encountered, and incrementally integrate them into an unfolding representation of the intended meaning, based on what they know about the words themselves, the structure of language and possible meanings. Importantly, in focusing on sentence level comprehension, the processes of speech perception and word recognition, which have also been investigated extensively using computational models (Magnuson et al., 2012), will be taken for granted. Similarly, models of how language is acquired by children (Alishahi, 2010) are not considered. For a comprehensive review of the numerous dimensions of psycholinguistic research see Spivey et al. (2012).

Perhaps the greatest challenge to developing theories and models of comprehension, as is the case for many areas of cognitive modelling, is that the central players – the nature of mental representations, the constraints that govern their construction, the processes involved in constructing representations, and how these processes interact – cannot be directly inspected using behavioral or neurophysiological methods. Furthermore, most online measures of human language comprehension – whether reading times, event-related potentials, or activations of brain regions – are known to be influenced by a range of factors, likely reflecting multiple underlying cognitive processes. It is therefore essential that explicit computational linking theories also be developed that identify precisely how cognitive processes are indexed by observable measures of comprehension. Only then can empirical data from a given measure be used correctly and consistently to inform computational theories that best reflect the nature of the human language comprehension system. For example, word by word reading times offer a behavioral measure of the time people spend on each word as they comprehend a sentence, which is generally taken to reflect cognitive effort (see Rayner, 1998). Increases in effort have then been associated with more specific mechanisms such as word recognition, lexical and syntactic disambiguation, reanalysis, as well as working memory. Neurophysiological measures such as event-related brain potentials, are scalp-recorded voltage fluctuations caused by post-synaptic neural activity, time locked to the onset of each word in a sentence. Observable components are generally taken to reflect the neural activity underlying specific computational operations carried out in given neuroanatomical networks. Of particular relevance to language comprehension are the N400 and P600 components (see Kutas and Federmeier, 2011, for a review). While there is some debate regarding precisely what cognitive processes these components index, the N400 is known to respond to semantically unexpected words, while the P600 has been demonstrated to be sensitive to more compositional syntactic, semantic and pragmatic violations. Other neurophysiological methods such as fMRI, offer further insight into the brain regions associated with particular linguistic features. Wehbe et al. (2014), for example show that machine learning methods can be used to predict activity in particular brain regions based on various lexical, syntactic and semantic features of words during reading of naturalistic texts. Importantly, however, our primary goal in computational psycholinguistics is not to model empirical measures as precisely as possible, but rather to develop models of language comprehension – that recover meaning from the linguistic signal – in a manner that is informed by, and consistent with, behavioural and neurophysiological measures.

Linguistic theories provide independently motivated accounts of the rules and representations that determine possible linguistic forms (syntax) and meanings (semantics). Indeed, all cognitive models must adopt some representational framework, minimally for defining the output of the system, but also possibly for intermediate levels of representation. Linguistic theories, however, traditionally emphasize human linguistic *competence* – formally characterizing "what" it means to know language – over the *performance* concerns about "how" the linguistic signal is encoded and decoded in real time. As a result, such accounts are often not entirely amenable to, nor informed by, demands of incremental processing, and are almost exclusively symbolic in nature, making them well-suited to more high-level symbolic processing models but more challenging to integrate naturally within neurocomputational accounts.

Another consideration that can help inform the development of computational theories of language is to consider broader theories of cognition. In particular, it has been argued that many cognitive systems can be viewed as *rational*, to the extent that they appear to behave in a manner that is *optimally adapted* to the *task* of that system and the *environment* in which it functions (Anderson, 1991). If a particular system – such as language comprehension – is regarded as being rational, then one can reason abstractly about "what" the optimal way to perform the task would be, what Marr (1982) refers to as a *computational* level theory. This in turn can be used to inform and constrain the development of suitable *algorithmic* level models that identify the actual mechanisms that instantiate the computational theory. Indeed, this approach has been dominant in computational psycholinguistics over the past two decades, resulting in the development of probabilistic theories that emphasize the role of likelihood in determining both human comprehension behavior in the face of ambiguity, as well as processing effort more generally.

Despite the vast empirical literature on language comprehension that has accrued over the last fifty years, advancements in linguistic theory, as well as the paradigm shift from symbolic toward probabilistic and subsymbolic (neural) computation, there is still relatively limited consensus regarding which computational mechanisms best characterize the comprehension system. There are several reasons for this: (1) the nature of mechanisms and representations is underdetermined by the empirical evidence, (2) experiments typically test binary predictions derived from hypotheses about some specific aspect of processing, disconnected from any complete model of comprehension (Newell, 1973), (3) interpretation of experimental findings is dependent on the linking hypothesis that is assumed, (4) results are often interpreted in isolation, and not reconciled with the broader literature. When developing and evaluating computational models of language, it is therefore important to take into account several dimensions:

Overarching behavior: People accurately understand the meaning of most utterances they encounter and do so highly incrementally, and typically without conscious difficulty. Models need to explain this generally accurate and effortless behavior, as well as pathological cases where people have difficulty.

Coverage: Models should not be tailored to individual phenomena and findings, but rather be consistent with as much relevant evidence as possible. As a consequence, it is important that models are in principle scalable with respect to their potential linguistic coverage. Further, models should ideally map into meaning representations, and explain interaction with world and situational knowledge, which are crucial for comprehension.

Linking hypothesis: While any given empirical measure underdetermines our characterization of the underlying comprehension mechanism, establishing accurate linking hypotheses to multiple complementary online measures – such as behavioral (e.g., reading times, visual attention), and neurophysiological (event-related potentials) – has the potential to mitigate this problem.

In this chapter a range of implemented computational theories of human sentence comprehension are reviewed, in the context of the above criteria, with a view to establishing both the points of consensus and important differences. A recently implemented neurobehavioral model of language comprehension, and its linking hypothesis with both neurophysiological and reading time measures, is then presented in greater detail in order to illustrate and integrate the concepts more concretely.

2. Early Perspectives on Sentence Processing

In pursuing the goal of characterizing the cognitive processes underlying the incremental, word-by-word nature of human sentence processing, early accounts focused on syntactic parsing: how is it that people integrate each word into a connected, semantically interpretable, yet possibly incomplete, analysis of the unfolding sentence (Frazier 1979)? While for most people brief introspection is enough to confirm this assumption of incrementality – that each word that is encountered contributes to furthering our understanding of the meaning being conveyed – this property has important consequences. Firstly, it formally constrains the set of mechanisms that one can consider with regard to the recovery of meaning, and secondly it entails that parsing mechanisms will need to make decisions in the face of substantial ambiguity. While much ambiguity in language – whether lexical, syntactic, and semantic – is eliminated by the end of the sentence, incrementality entails that decisions about how to integrate each word into the unfolding sentence interpretation must be made as soon as that word is encountered. This predicts that, if a decision taken at some point of ambiguity in the sentence is subsequently disconfirmed, it will be necessary for the parsing mechanisms to re-process the sentence, or restructure the current analysis, to accommodate the disconfirming word. This reprocessing cost is postulated to result in observable processing effort as manifested by, for example, word-by-word reading times. A classic illustration of this comes from the reduced relative clause ambiguity (Bever, 1970) in (1a) compared to its unreduced, and unambiguous, counterpart in (1b) (adapted from Rayner et al.,1983):

(1a) “The florist sent the flowers smiled.”

(1b) “The florist who was sent the flowers smiled.”

When “sent” is first encountered in (1a) it is in fact ambiguous as being either a simple past verb, or a past participle. As illustrated in Figure 1 (ignore the probabilities for now), the parser must therefore decide whether to analyze it as the main verb of the sentence (and thus as simple past) as shown in the first parse tree, or as a past participle which begins a (reduced) relative clause, illustrated by the second tree. Frazier (1979) argued that human preferences for a range of such local structural ambiguities could be explained by two simple decision principles. The Minimal Attachment (MA) principle postulated that the parser should prefer less complex syntactic analyses (i.e., fewest nodes in the parse tree). In this case, MA predicts that “sent” is initially analyzed as the main verb – as this parse tree has fewer nodes when “sent” is processed. While the noun phrase “the flowers” is consistent with either analysis, the verb “smiled” – which can only be the main verb – disconfirms the previously adopted main clause analysis, explaining why most people find (1a) to be difficult: once “smiled” is encountered either substantial reprocessing effort is required to construct the reduced relative clause interpretation of “sent the flowers”, or it simply cannot be integrated at all.

In cases where two possible analyses are equally minimal according to MA, a second principle – Late Closure – postulates that the word should be attached to the most recently built part of the parse tree. This commonly occurs with modifying phrases which can be associated with several phrases, as in (2):

(2) “Someone shot the governor of the company that had been sold/elected”

Here, when “that” is reached, the parser must begin the construction of a relative clause that can modify either “governor” or “company”. LC predicts this will be attached to “company”, explaining why less reading effort is observed when the final word of the relative clause is consistent with that attachment (e.g. “sold”), compared to when it forces attachment to “governor” (e.g. “elected”).

Beyond these two decision principles for resolving structural ambiguity, Frazier explicitly assumes several other important characteristics of the parsing mechanism. Firstly, the parser is strictly serial, in that once a preferred parse has been constructed, alternatives are no longer considered. Second, the parser is purely syntactic, with access to only basic part of speech information about incoming words of the sentence, and decision strategies that are determined by structural properties of the parse. The parsing model can thus be viewed as very strictly modular, in the sense of Fodor (1983), in that the initial incremental parsing and disambiguation process has no access to, and is not influenced by, either detailed lexical (e.g., meaning, frequency, subcategorization) or semantic (e.g., plausibility) information. The overriding motivation for this collection of assumptions is that they contributed to reducing the amount of computation, and thus effort, involved in achieving real time comprehension. Importantly, however, no implementation or precise parsing algorithm is provided, and it is interesting that early attempts to implement Frazier's decision strategies reveal that it is not straightforward. Firstly, strictly incremental parsing algorithms are in fact not straightforward to implement for conventional (hierarchical) phrase structure grammars (see Crocker, 1999, for discussion). Indeed, Marcus (1980) developed a deterministic model of human parsing that required extensive non-incremental "look ahead" capabilities to explain why many structural ambiguities do not cause substantial processing difficulty, but in doing so completely violates any notion of incrementality. Secondly, MA entails that the parser be able to compare competing alternative parses with respect to their global structural properties, something which cannot be accomplished fully in terms of standard serial parsing operations (Pereira, 1985), emphasizing the importance of developing fully specified computational theories, rather than relying on purely verbal formalizations.

Following in Frazier's footsteps, however, several computational models of sentence processing were developed that, while differing in various important respects, shared her view of a serial, incremental, and largely modular syntactic processing architecture. For example, Crocker (1996), building on proposals by Pritchett (1988), implemented a model for English and

German which prioritized thematic role assignment (agent, theme, location, etc.) over simple structural decision principles like MA. Stevenson (1994) proposed a related hybrid network model of human parsing and disambiguation which emphasized both role assignment and minimal structure building. Gibson (1998), in contrast, developed a model in which memory load (unresolved dependencies) and locality contribute to determining comprehension effort as well as preferences in resolving ambiguity, while Lewis and Vasishth (2005) use the ACT-R framework (see the chapter on Cognitive Architectures by Niels Taatgen and John Anderson in this handbook) to model the role of memory retrieval in determining parsing difficulty.

It is worth noting that all of these models assume that the human parser incrementally recovers a grammatically licensed and semantically correct interpretation of a sentence. There are, however, a range of findings suggesting that comprehenders – depending on their goals, and situational demands – may not always analyse sentences fully or even correctly (Sanford & Sturt, 2002; Ferreira, 2003). Global attachment ambiguities as in (2) for example, might simply be left unresolved if the final verb does not disambiguate the two readings (e.g., "discredited" instead of "sold"). Other evidence suggests that people sometimes misunderstand simple Noun-Verb-Noun sequences "the dog was bitten by the man", failing to recover the passive reading, and rather using an Agent-Action-Patient heuristic (i.e. "the dog bit the man"), particularly when that reading is more plausible (Townsend & Bever, 2001; Ferreira, 2003; Gibson et al., 2013). Studies using event-related potentials, have also been taken as providing evidence that anomalous sentences like "After an air crash, where should the survivors be buried?" (where the survivors are presumably still alive) elicit no effect in the N400 component, despite this component often being found for semantically unexpected words (Sanford et al., 2011). Related evidence from role-reversal anomalies, such as "The hearty meal was devouring …" also elicit no N400 effect (Kim & Osterhout 2005; see also Hoeks et al., 2004; van Herten et al. 2005), suggesting that people may construe a plausible meaning ("the meal was devoured") for an implausible sentence (though an alternative interpretation is considered in section 5). Taken together, evidence that comprehenders may modulate the depth and veracity of linguistic processing – possibly as a function of their communicative goals and available cognitive resources, but also their prior knowledge and expectations – has been used to argue that the comprehension system may in some circumstances function in a manner that is "good enough" (Ferreira et al., 2002). The diversity of these phenomena – spanning lexical meaning, grammaticality, semantic role reversals, as well as the underspecification of syntactic ambiguity – has thus far eschewed any uniform treatment (see Ferreira & Patson, 2007, for discussion), though models have been proposed which address particular phenomena such as role-reversal anomalies (e.g. Gibson et al., 2017; Rabovsky & McClelland, 2019). In general, the focus of discussion here will be on modelling the process of full understanding that has been attested in many psycholinguistic studies. Nonetheless, better understanding of how the human comprehension system modulates its depth and accuracy of processing may offer important insights into the nature of the mechanisms and representations involved.

3. Probabilistic Models and Rational Approaches

Many of the models discussed above, and particularly that of Frazier, assumed that cognitive limitations – coupled with the time sensitive demands of real time comprehension – were central in shaping the nature of the human parsing mechanism. That is, the need to quickly and incrementally structure the incoming signal into an interpretable representation is used to motivate serial processing (rather than constructing multiple analyses in parallel) and simple modular decision principles. This perspective was fundamentally challenged by increasing empirical evidence that a variety of non-syntactic factors – such as prior experience (frequency), plausibility, context, and world knowledge – can rapidly influence disambiguation (MacDonald et al., 1994). This resulted in a shift away from highly restricted "serial, syntax first" models, toward models that start with the assumption that people bring considerable computing resources and diverse relevant information sources to bear on language comprehension. With hindsight, this perspective can be seen as a shift away from viewing comprehension as a system shaped by *limitations* on cognitive processes, to one in which it is viewed as more *rational* (Anderson, 1991). That is, a view in which the behavior of the comprehension system is optimally adapted to the task (obtaining the correct meaning), given the environment (an incremental and ambiguous signal). Importantly, this view invites theorists and modelers to constrain our search for the specific mechanisms underlying comprehension, by first thinking carefully about what the goal of the system is, and how it can be optimally achieved:

> *An algorithm is likely understood more readily by understanding the nature of the problem being solved than by examining the mechanism (...) in which it is solved.* (Marr, 1982, p.27)

Taking this view, one can theorize what the goal of rational comprehension system might be (Marr's *computational* level), and then consider what cognitively plausible mechanisms and representations (Marr's *algorithmic* level) might instantiate such a theory (Crocker, 2005). As a first approximation, it seems reasonable to suggest that the comprehension system's goal is to recover the most likely interpretation of the input, which can be formalized as in equation 1.

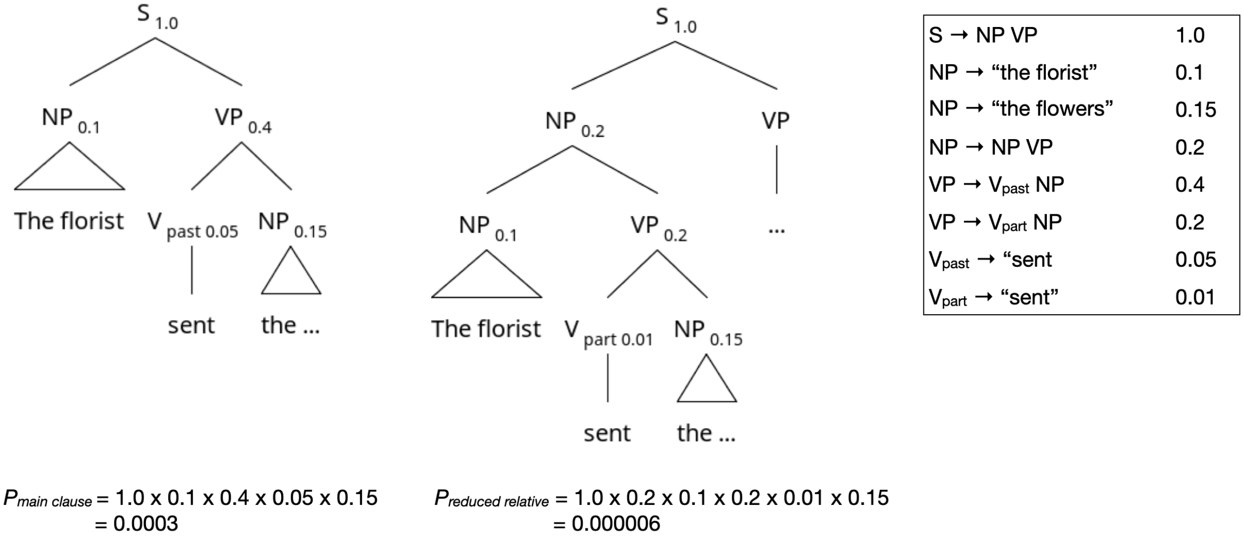
$$\text{(Eq 1) } \hat{I} = \text{argmax } P(i|s,K)$$

where i ranges over the possible interpretations of the sentence s. That is, this function states that "what" the comprehension system does is seek to identify the interpretation $\hat{I}$ that has the highest likelihood given the sentence itself, and our relevant knowledge K. Given the assumption of incrementality, this formalization can also be extended to the word-by-word construction of sentence meaning (but see Chater et al., 1998, for an alternative rational analysis).

With this computational level theory in place, one can now consider what algorithmic level models are plausible instantiations of this theory. Jurafsky (1996) built on the abundant evidence for the role of frequency in both lexical and syntactic disambiguation to motivate a probabilistic likelihood-based model of the parsing process. The model departs from Frazier in two key respects: (a) it constructs all possible alternatives in parallel, and (b) it determines their probability on the basis of lexical and syntactic frequency information, including probabilistic information about the subcategorization preferences of individual verbs. Returning to the example (1a) above, Figure 1 illustrates how a probabilistic grammar and lexicon,[1] shown in the panel on the right, can be used to assign a probability to each possible parse of the sentence after the verb "sent" is encountered. That probability is simply the product of the probabilities of each rule used to construct the parse tree. In this case, the reduced relative clause is assigned a probability more than two orders of magnitude lower than the main clause. This is due to a number of reasons: the additional relative clause rule (NP → NP VP), the lower probability of the

[1] This grammar is highly simplified for expository purposes. Phrases such as "the florist" should of course be fully parsed, the analysis assigned to the reduced relative clause is simplified, and the probabilities themselves were constructed to reflect relative likelihoods of the two structures. Typically the grammar and the probabilities would be determined using a large parsed corpus.

reduced relative (VP) itself, and the lower probability for “sent” to be a past participle. To explain why sentences such as this (and many others) are difficult, Jurafsky proposes a linking hypothesis under which parses that are too low compared to the highest probability parse – as is the case in this example – are eliminated by the parser (or “pruned”), such that they can’t be retrieved later in the sentence. While in this case the model makes the same prediction as Frazier, the model explains other instances of this ambiguity which do not cause the same degree of difficulty, due to differences in the specific probabilities. That is, the likelihood of the reduced relative clause analysis may be sufficiently close to that of the main clause parse, that it is not pruned, meaning that when the main verb “smiled” is encountered, the relative clause analysis is still available.



S → NP VP	1.0
NP → “the florist”	0.1
NP → “the flowers”	0.15
NP → NP VP	0.2
VP → V_{past} NP	0.4
VP → V_{part} NP	0.2
V_{past} → “sent	0.05
V_{part} → “sent”	0.01

$P_{main\ clause} = 1.0 \times 0.1 \times 0.4 \times 0.05 \times 0.15$
$= 0.0003$

$P_{reduced\ relative} = 1.0 \times 0.2 \times 0.1 \times 0.2 \times 0.01 \times 0.15$
$= 0.000006$

Figure 1: Syntactic analyses of the main clause (left) and reduced relative clause (middle) ambiguity. A probabilistic context-free grammar (right) used to derive the probabilities of both parse trees.

As an algorithmic level account, Jurafsky proposed a concrete mechanism that can be seen as approximating the computational level theory. Specifically, an incremental parallel parser exploits a probabilistic context-free grammar to approximate the true probability of syntactic analyses and substitutes bounded parallelism (implemented using beam search based on the pruning of low probability parses) in place of full parallelism (which is often viewed as cognitively implausible). Crocker and Brants (2000) propose a wide-coverage probabilistic parsing model which can be seen as an alternative algorithmic level instantiation of the same computational level likelihood theory, differing primarily in the role of verb-frame information, the nature of the pruning mechanism, and the proposal of a linking hypothesis in which reranking of (non-pruned) parses is assumed to result in more graded increases in reading difficulty.

One limitation of these models, however, is the lack of any means to assess semantic plausibility of competing syntactic analyses. For example, contrasting (1a) with (1c), people typically find the (1c) version easier to understand, as evidenced by reduced total reading times (Rayner, 1983). This is because while “florist” is a good Agent of the “send flowers” event, encouraging the comprehender to prefer a main clause analysis, “performer” is a better Recipient of such an event, rendering the relative clause interpretation easier to recover.

(1c) “The performer sent the flowers smiled.”

The increasing amount of empirical evidence from reading times demonstrating the rapid influence of semantic knowledge on human disambiguation (e.g., Trueswell et al., 1994) motivated several non-modular “constraint-based” theories. These accounts posit that probabilistic constraints at all relevant levels (from phonology and morphology through to syntax, semantics and constraints provided by the context) contribute directly and immediately to determining which interpretation — among the possible grammatical alternatives — is best (Macdonald et al., 1994; Tanenhaus et al., 1995). This approach was instantiated in the Competition-Integration Model (CIM), illustrated in Figure 2, in which competing interpretations (I_1 and I_2) are simultaneously represented,[2] with their relative activation being determined by a collection of probabilistic constraints, each providing more or less support for a particular interpretation, and with each constraint having its own weight compared to the other constraint (McRae et al., 1998). Crucially these constraints can in principle reflect any relevant source of information, such as lexical frequency bias and semantic plausibility, or even broader contextual constraints, resulting in their immediate influence on the resolution of ambiguity.

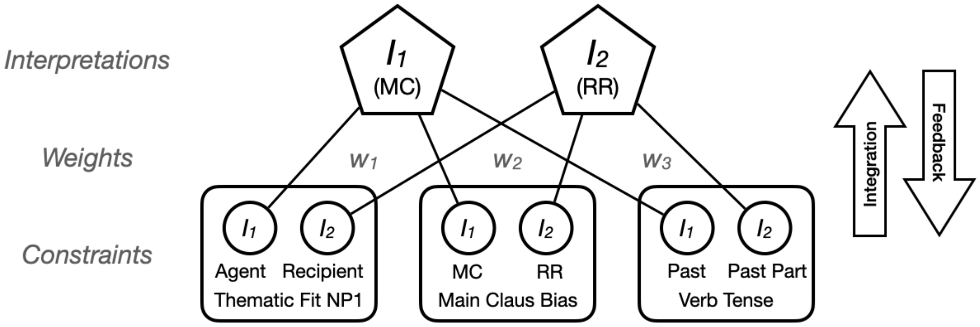


Figure 2: The Competition-Integration Model (McRae et al., 1998).

Constraint biases in the model are established independently using corpus frequencies (e.g., to determine the main clause versus relative clause frequency, and frequency of “sent” as either

[2] In contrast with other models discussed here, the CIM does not include any mechanism to construct the alternative interpretations, but rather models how the comprehension system resolves the ambiguity.

simple past or past participle) or human judgement studies (e.g., to determine how likely a "florist" or "performer" is to either send or receive flowers). To model the online disambiguation process, the probability of the two possible interpretations is computed as the weighted average of the probability assigned to it by each of the individual constraints (the "integration" step). A recurrent "feedback" mechanism readjusts the constraint biases to reflect the interpretation activations. The model then continues these integration-feedback cycles until one of the interpretations reaches threshold. This number of cycles is postulated to quantitatively index disambiguation effort that is reflected in reading times. As McRae et al. (1998) demonstrate, this approach is able to capture the influence, particularly of the thematic fit of the initial noun as either an Agent or Recipient of "sent", on modulating the difficulty of these ambiguities. While this can be seen as another algorithmic instantiation of a likelihood model, it differs significantly with regard to the proposed linking hypothesis, which is only indirectly determined by the likelihoods of various constraints. One limitation of this approach, however, is that a new model must be constructed to model each ambiguous construction type and the constraints relevant to it (see Tanenhaus et al., 2000, for an overview). To address this shortcoming, Pado et al. (2009) demonstrated how thematic fit could be estimated from large corpora and integrated into a broad-coverage incremental probabilistic parsing architecture similar to those discussed above, while also retaining a likelihood-based reranking linking hypothesis similar to Crocker and Brants (2000).

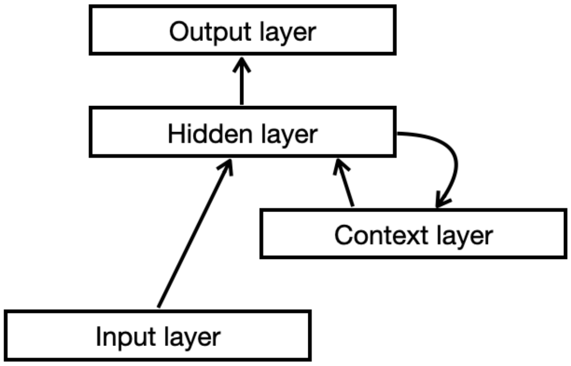


Figure 3: The Simple Recurrent Network architecture (Elman, 1990).

In addition to the symbolic implementations of likelihood models above, many connectionist models also maximize the likelihood of their output representation, for a particular input, as a consequence of learning algorithms that minimize error on average (Rumelhart et al., 1986), and thus reflect the statistical properties of their training environment (see the chapter on Connectionist Models of Cognition by Michael Thomas and James McClelland in this handbook). One particularly well-known illustration of this comes from Elman's (1990) Simple Recurrent neural Network (SRN) – a three-layer feedforward network: **input** ↦ **hidden** ↦ **output**, in which at a timestep *t*, the **hidden** layer receives additional input from a **context** layer that contains the activation pattern of the **hidden** layer at timestep *t-1*. Crucially, the context layer provides a memory of words that have been processed previously, enabling the network to draw upon the entire unfolding sentence despite processing on a word-by-word basis (Figure 3). Elman (1990) demonstrated that, when trained to predict the next word – a task which is inherently non-deterministic for human languages – the model's output was strongly correlated with the conditional n-gram likelihoods determined from the training corpus. While modeling next word prediction is clearly not equivalent to language comprehension, the SRN architecture has also been used to map into fixed "sentence gestalt" representations of sentence-level meaning, which represent the main action and its associate role-fillers (McClelland et al., 1989) that also reflect likelihood (Mayberry et al., 2009; Brouwer et al., 2017; Rabovsky et al., 2018). However, while next-word-prediction networks can easily be scaled to unrestricted language (see e.g., Aurnhammer and Frank, 2019), a long-standing concern regarding connectionist models of human comprehension relates to the scalability and linguistic adequacy of such thematic role and sentence gestalt representations for recursive sentence structures, as well as to whether a fixed number of output units can represent the full compositional and hierarchical nature of possible meanings (see Lopopolo & Rabovsky, 2021, for one approach). Recent advances in connectionist computational linguistics have begun to address these issues (see, e.g., Bowman et al., 2016, and Linzen & Baroni, 2021), potentially offering solutions which may be relevant to future cognitive models.

4. Expectation-based Models of Sentence Comprehension

The probabilistic models outlined above can be viewed as algorithmic level instantiations of a computational theory which seeks to maximize the likelihood of recovering the correct interpretation. However, while all models have in common a linking hypothesis for which high likelihood interpretations will be easier to process than low likelihood ones, they differ substantially with regard to why this is the case – is it because low probability interpretations are completely pruned, or simply ranked lower, or is it due to the increased cycles of competition needed for a low probability interpretation to reach a threshold. That is, having identified a computational theory of the comprehension system, the linking hypotheses are only stated at the algorithmic level. Hale (2001) builds on consensus around likelihood-based architectures in proposing a computational level linking theory, which is grounded in Shannon's (1948)

Information Theory. Specifically, Information Theory provides a mathematical framework to determine the amount of information that is conveyed by an event (such as encountering a word w_i) – also known as its *surprisal* – as determined by its expectedness in a given context (since the likelihood of a word is heavily influenced by the context in which it appears):

$$\text{(Eq 2) } surprisal(w_i) \ = \ -log_2 P(w_i|Context)$$

Hale proposes that the effort required to process each word of an unfolding sentence should be proportional to its surprisal – the number of bits of information each word conveys, given the context in which it occurs (such as the preceding words). The prediction is clear: words that are highly expected in a particular context convey little information, and require little effort to process, while words that are unlikely convey more information, and entail more effort. At face value, this claim appears almost trivial. It has long been known, for example, that a word's *Cloze* probability (Taylor, 1953) – namely, the likelihood that people will complete a particular context with a given word – is a robust predictor of its reading time and skipping probability (Rayner & Well, 1996), as well as N400 amplitude (Kutas & Federmeier, 2011). For this reason, many psycholinguistic experiments determine, and control for, the Cloze probability of critical words. However, as outlined by Hale (2001), and later expanded upon by Levy (2008), surprisal can also be seen as quantifying the reranking cost of an incremental probabilistic parser. Specifically, the surprisal of a word w_i is determined based on prefix probabilities determined by the parser, based on the first *i* words of the sentence:

$$\text{(Eq 3) } surprisal(w_i) = \ -log_2 P(w_i|w_1...w_{i-1}) \ = \ -log_2 \frac{P(w_1...w_i)}{P(w_1...w_{i-1})}$$

$$= -log_2 \frac{\sum_{T \in Trees} P(T|w_1 \dots w_i)}{\sum_{T \in Trees} P(T|w_1 \dots w_{i-1})}$$

Under this formulation, the surprisal of a word w_i is determined from the probability of the sentence prefix up to an including w_i – which is the sum of the probabilities of all the parses that span that prefix – divided by the probability of the sentence prefix before w_i was encountered. Reconsidering the reduced relative clause, the appearance of "smiled" means that, while the denominator in Equation 3 will be the sum of both possible analyses shown in Figure 1, the numerator will only include the very low probability relative clause parse, resulting in high surprisal. One way to view surprisal, therefore, is as reflecting the relative loss of probability mass associated with the much more likely main clause parse, and the need to "shift attention" toward a much lower probability alternative. Indeed, Levy (2008) points out that surprisal at word w_i is equivalent to the change in the probability distribution over all possible parses after w_i is encountered compared to before it was encountered, as quantified by Kullback-Leibler divergence (KLD). As such, surprisal can be viewed as characterizing the effort associated with reranking, or shifting of attention, among possible parses based on the integration of word w_i.

Importantly, however, quantifying the effort of resolving such structural ambiguity is simply a special case of surprisal, which can just as well explain difficulty in unambiguous constructions, such as the preference for subject relative clauses over object relative clauses (Hale, 2001), or simply a word's expectancy in a particular context. That is, even if only a single parse is possible up to word w_{i-1}, not all continuations will be equally likely as there is still *uncertainty* regarding how the sentence will unfold lexically and syntactically, both of which will influence the unfolding probability of the utterance at $w_{1 \dots i}$, compared to $w_{1 \dots i-1}$ and thus the surprisal induced by w_i (see Roark et al., 2009, for details). Furthermore, as Equation 2 makes clear, while Surprisal theory can be implemented in terms of parse probabilities, there are many other algorithms that can determine the likelihood of a word in context, including so-called *language models*, which don't recover any interpretation at all, such as statistical n-gram models, and connectionist word-prediction models based on SRNs and LSTMs (Aurnhammer & Frank, 2019; Michaelov & Bergen, 2020). Indeed, this highlights what Levy (2008, pp. 1132-1133) refers to as a *causal bottleneck*, namely that "many different classes of generative stochastic process can determine conditional word probabilities" and will thus similarly account for empirically observed surprisal effects.

The empirical coverage of surprisal theory is considerable in explaining reading-time behavior observed for many ambiguities (Hale, 2001; Levy, 2008) and in more naturalistic texts (Boston et al., 2008; Demberg & Keller, 2008; Smith & Levy, 2013). Similarly, surprisal has been found to correlate with neurophysiological measures, typically the N400, in both controlled (Delogu et al., 2017; Michaelov & Bergen, 2020; Staudte et al., 2021) and naturalistic (Frank et al., 2015; Brennan & Hale, 2019) studies. However, as the causal bottleneck illustrates, even relatively uninteresting language models can capture these effects. Thus, in the context of building models of language comprehension, the goal must rather be to explain why surprisal correlates with cognitive effort as a consequence of the mechanisms and representations that underlie the comprehension process, and fully characterize how surprisal is determined in such a mechanism, and manifest across the spectrum of relevant observable measures.

While the instantiation of surprisal as KLD over syntactic analyses elevates Surprisal Theory as an overarching, explanatory linking theory of word-by-word processing difficulty, syntactic analyses are still at best a proxy for utterance interpretations. That is, comprehension is not about deriving a structural analysis of a sentence per se, but about recovering a 'situation model'-like representation of utterance meaning, which may also go well beyond the literal propositional content conveyed by an utterance (Johnson-Laird, 1983; Van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998). For instance, understanding a simple sentence such as "John is sleeping.", presumably does not just involve extracting the proposition `sleep(john)`, but may also include the 'world-knowledge'-driven inferences such as `wear(john,pyjamas)`, `in(john,bed)`, and `time_of_day(night)`. Moreover, accumulating evidence shows that world knowledge affects word processing difficulty above and beyond linguistic experience alone (see Warren & Dickey, 2021, Venhuizen et al., 2019, and the references therein). Hence, for Surprisal Theory to scale up, models should be developed in which comprehension involves recovering rich 'situation model'-like utterance meaning representations capturing 'world-knowledge'-driven inferences, rather than deriving syntactic analyses alone, and online processing in these models should be sensitive not only to the likelihood of syntactic analyses based on linguistic experience, but also to the likelihood of utterance meanings, based on knowledge about the world.

Venhuizen, Crocker, and Brouwer (2019) have recently proposed such a model of "comprehension-centric" Surprisal. Their model is a three-layer Simple Recurrent neural Network (recall Figure 3) that processes sentences on a word-by-word basis, and incrementally recovers a 'situation model'-like utterance meaning representation (Frank et al., 2003, 2009; Venhuizen et al., 2019, 2021). The building blocks for these meanings representations are vectors for atomic propositions, such as `sleep(john)` and `walk(mary)`, which can be combined into vectors of propositions of arbitrary complexity using logical operations. That is, the meaning of atomic and complex propositions is defined relative to a number of observations of *states of affairs* in the world, in which a proposition is either true or false. Indeed, propositional meaning is thus defined in terms of co-occurrence relative to these observations of *states of affairs*, which serve as cues towards determining the truth-conditions of a logical expression. This approach is analogical to how linguistic contexts offer cues for determining lexical meaning in distributional lexical semantics (see Lenci, 2018, for a review), but importantly supports the representation of arbitrarily complex compositional meanings. These meaning representations are inherently probabilistic as well. That is, as the number of observations relative to which the meaning of a proposition is defined grows, the fraction of observations in which the proposition is true increasingly approximates its probability in the world. Given the logical nature of these representations, the probability of two propositions co-occurring, as well as the conditional probability between propositions directly derives from the vector representations, thereby allowing for 'world knowledge'-driven inferences (see Venhuizen et al., 2021, for details).

Taken together, a finite set of atomic propositions, and a finite set of observations that describe the state of each of these propositions in terms of their truth or falsehood, thus define a meaning space that is inherently probabilistic, which allows for "world knowledge"-driven inferences and the compositional derivation of complex propositions (see Venhuizen et al., 2021 for a detailed exposition). Ideally, this meaning space should capture the structure of the world in terms of hard co-occurrence constraints (e.g., certain propositions can be true at the same time) as well probabilistic co-occurrence constraints (e.g., certain propositions are more likely to co-occur than others). To illustrate this, Venhuizen and colleagues constructed such a meaning space by deriving 150 observations from a high-level description of the world, covering forty-five atomic propositions pertaining to activities on a night out on the town: e.g., `enter(beth,cinema)`, `order(beth,popcorn)`, `enter(thom,restaurant)`, and so forth. They then trained their SRN to map sentences, on a word-by-word basis, onto their corresponding sentence-final meaning representation, that is, a vector representing atomic or complex propositional meaning. As certain sentence-prefixes may overlap (e.g., "thom entered [bar/restaurant]"), the model will produce vectors at sentence-intermediate words that lie at the crossroads of potential sentence-final meanings. In other words, comprehension in the model is effectively word-by-word navigation through meaning space. Figure 4 visualizes this comprehension as meaning space navigation process in three-dimensional space (using multi-dimensional scaling). Given the sentence-initial word "thom", the model moves towards a (colored) point in space that is in between all potential sentence-final meanings (gray points). Upon encountering the next word "ordered", the model then moves in the direction of sentence-final meanings pertaining to `order(thom, [...])`, and so forth, until the sentence-final word is reached upon which sentence-final meaning will be recovered.

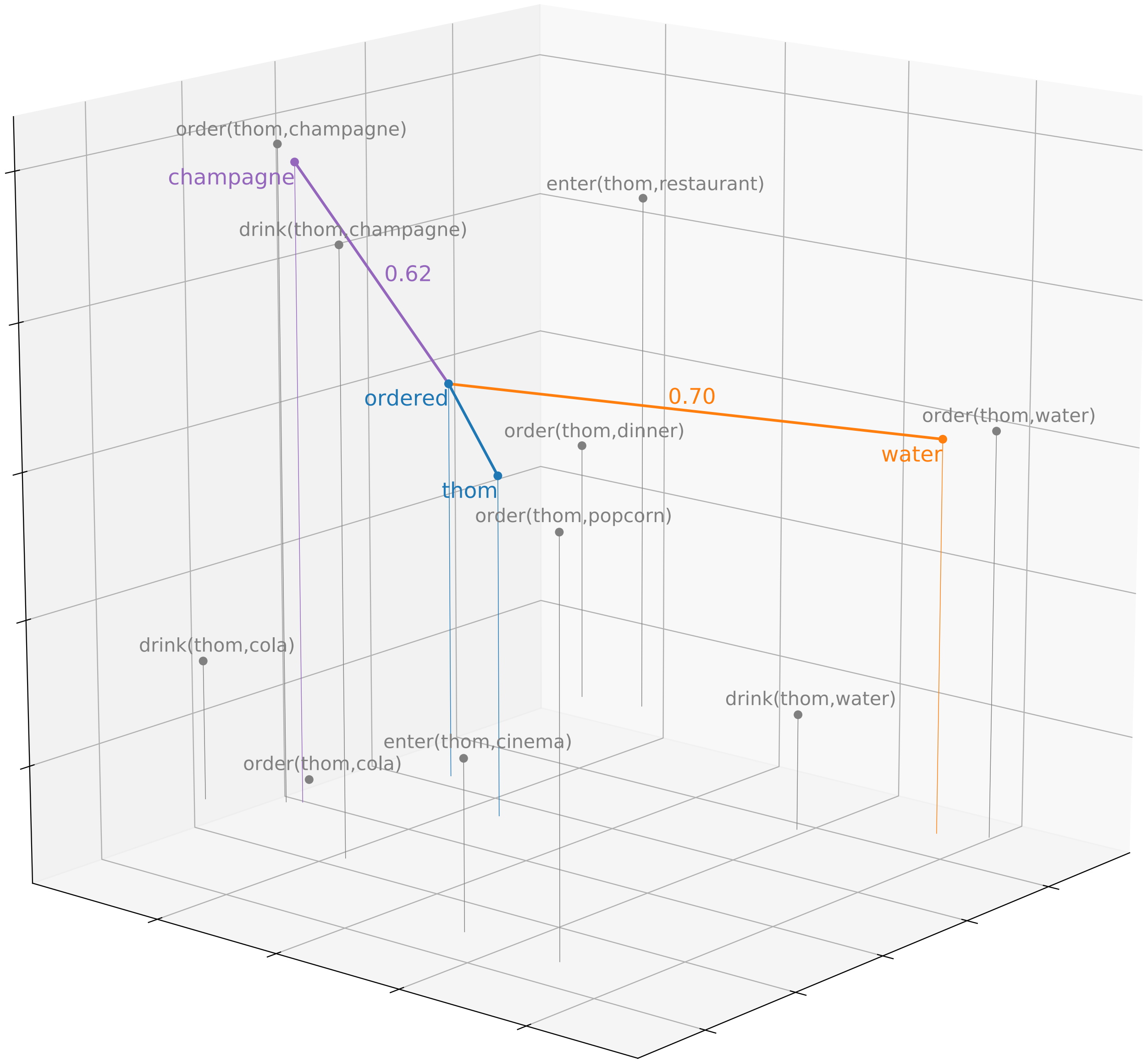


Figure 4: Three-dimensional visualization of comprehension as meaning space navigation.

Crucially, as can be seen in Figure 4, certain words trigger larger movements through meaning space than others (compare “water” to “champagne”). In general, words that trigger larger movements can be thought of as inducing a less expected shift in meaning than words that trigger smaller movements; that is, prior to processing a next word, the model will be in a state that is closer to more expected continuations than to more unexpected ones. Venhuizen et al. harness the probabilistic nature of the meaning space to quantify this 'comprehension-centric' notion of expectancy by defining the surprisal induced by a word w_t as the negative log probability of the meaning as constructed by the model after processing w_t, the activation pattern produced at the **output** layer of the model at *t*, given the meaning prior to encountering word w_t, the activation pattern produced at the **output** layer at *t-1*:

$$\text{(Eq 4) } surprisal(w_t) \; = \; -logP(output_t \mid output_{t-1})$$

The numbers in Figure 4 show that given the interpretation constructed after processing the sentence-initial fragment "thom ordered [...]", 'comprehension-centric' surprisal is higher for "water" (.70) than for "champagne" (.62). Note that while this notion of 'comprehension-centric' surprisal is indeed closely related to distance in meaning space, they are not strictly the same thing mathematically. Finally, as the model navigates the meaning space on a word-by-word basis, this notion of 'comprehension-centric' surprisal is effectively similar to KLD over syntactic analyses; the processing of a word potentially prunes away sentence-final meanings, and the more probability mass is pruned a way, the higher surprisal. Importantly, Venhuizen et al. (2019) trained the model such that it is exposed to certain sentences more frequently than other sentences (linguistic experience), and such that certain sentences are mapped onto meanings that are more probable in the meaning space than other meanings (world knowledge). Their simulations demonstrate that the model combines cues from these information sources, balancing them according to their relative strengths.

To illustrate, consider the following ‘comprehension-centric’ surprisal-based account of the reduced relative clause ambiguity (1a), which induces increased processing effort relative to its unreduced counterpart (1b). Prior to encountering the disambiguating main verb “smiled”, sentence (1a) is compatible with a meaning in which either the proposition `send(florist,flowers)` or in which `receive(florist,flowers)` is inferred. Here, world knowledge will dictate that florists are more likely to send rather than receive flowers, thereby biasing towards the interpretation in which `send(florist,flowers)` holds. Linguistic experience, in turn, may support this bias, as noun phrases in subject position are most often followed by the main verb, and tend to be agents of that verb. Upon encountering the main verb “smiled”, however, this bias is disconfirmed, and the model needs to rule out the proposition `send(florist,flowers)` by inferring its negation `!send(florist,flowers)`, and draw the inferences that `receive(florist,flowers)` and that `smile(florist)`. Indeed, this involves pruning away a high probable point in meaning space and moving towards a less probable one. Crucially, this is not necessary in (1b) where the relative pronoun “who” informs the model to adopt such a relative clause analysis immediately. Hence, depending on whether or not the model has encountered the relative pronoun “who”, it will find itself in a different place in meaning space prior to encountering the main verb “smiled”. Consequently, the surprisal induced by “smiled” in (1a) will be higher than in (1b), because in (1a) it will trigger a less likely transition in meaning space than in (1b). This increased processing cost at “smiled” in (1a) will be reduced, however, if the sentence-initial noun phrase is replaced with a good recipient for “the flowers”, like “the performer” (1c), since `receive(performer,flowers)` will be much more likely than `receive(florist,flowers)`. In this case world knowledge will bias the model towards a reduced relative interpretation, even when no explicit relative pronoun is present.

In sum, word-by-word, incremental sentence comprehension in the Venhuizen et al. model can be conceptualized as word-by-word meaning space navigation. Both the linguistic experience of the model and the world knowledge contained within the meaning space, influence how precisely the model traverses the meaning space during processing. At any point in processing, more frequent sentence continuations (linguistic experience) and more probable meanings (world knowledge) are favored over less frequent sentence continuations and improbable meanings. Moreover, if linguistic experience and world knowledge are in conflict, their relative weightings will determine model behavior. Surprisal in the model is ‘comprehension-centric’ and derives directly from the probabilistic meaning representations that the model constructs. Higher surprisal ensues when an incoming word induces a less expected change in utterance meaning, while a more likely change leads to lower surprisal.

5. A Neurobehavioural Model of Sentence Comprehension

The ‘comprehension-centric’ notion of surprisal proposed by Venhuizen et al. (2019) predicts that processing cost is directly related to the word-by-word updating of an unfolding utterance interpretation. An open question, however, is how this processing cost is reflected in the different neurophysiological and behavioral indices of processing difficulty, in particular the N400 component and the P600 component of the ERP signal, as well reading times (henceforth RTs). Brouwer, Delogu, Venhuizen, and Crocker (2021) have recently proposed an explicit neurocomputational model that addresses this question. The core of this model is a neurocomputational instantiation of the Retrieval-Integration account of the N400 and the P600 in language comprehension (Brouwer et al., 2017).

On the Retrieval-Integration account, the N400 component of the ERP signal – a negative deflection that reaches maximum amplitude at around 400ms post word onset – reflects the retrieval of the meaning of an incoming word from long-term memory. This retrieval is facilitated, leading to a reduction in N400 amplitude, when word meaning is primed by lexical and/or contextual cues; for instance, continuing “He spread his warm bread with [...]” with “socks” leads to a larger N400 than when continuing it with “butter” (Kutas & Hillyard, 1980), as the latter is primed to a larger degree than the former. In turn, the P600 component of the ERP signal – a positive deflection reaching maximum around 600-800ms post word onset – indexes the integration of retrieved word meaning into the unfolding utterance interpretation. P600 amplitude increases whenever the meaning of an incoming word incurs structural, semantic, or pragmatic integration difficulty. The Retrieval-Integration account thus predicts word-by-word processing to proceed in retrieval (N400) and integration (P600) cycles, such that in addition to an N400-effect, for the contrast “He spread his warm bread with socks/butter”, a P600 effect is also predicted indicating difficulty in integrating the meaning of socks into the unfolding utterance interpretation. This account contrasts with models in which the N400 is assumed to also index integration processes (Baggio & Hagoort, 2011; Rabovsky et al., 2018), as well as accounts that link the P600 to syntactic (e.g., Gouvea et al., 2010) or more general conflict resolution processes (Rabovsky & McClelland, 2019).

The Retrieval-Integration account makes the prediction that implausible words may nonetheless be highly associated with the sentence context, facilitating retrieval and attenuating the N400, while integration difficulty is still predicted to be reflected in the P600. This is precisely the case with the role-reversal example discussed in Section 2, where continuing “The hearty meal was [...]” with “devouring” does not elicit an N400, as “devouring” and “devoured” are equally primed by the context, but rather produces a larger P600 than continuing it with “devoured” (Kim & Osterhout, 2005), as according to our linguistic and world knowledge “the hearty meal” is a poor agent for devouring (but is a good patient of “was devoured”). How other models explain such findings is considered below.

The original neurocomputational model instantiating such Retrieval-Integration cycles was shown to account for key semantic processing phenomena such as those above, but was somewhat limited in coverage due to its use of linguistically impoverished “thematic role”-based utterance meaning representations (as discussed previously). In a more recent instantiation of the model, Brouwer et al. (2021) replaced these “thematic-role”-based meaning representations with the “situation model”-like meaning representations introduced above (Venhuizen et al., 2019). The resultant comprehension model recovers “situation model”-like utterance meaning interpretations on a word-by-word basis, and produces estimates of the N400, reflecting the effort involved in retrieving word meaning, the P600, indexing the work involved in integrating the retrieved word meaning into the unfolding utterance interpretation, as well as of surprisal, reflecting the likelihood of the change in utterance meaning induced by a word.

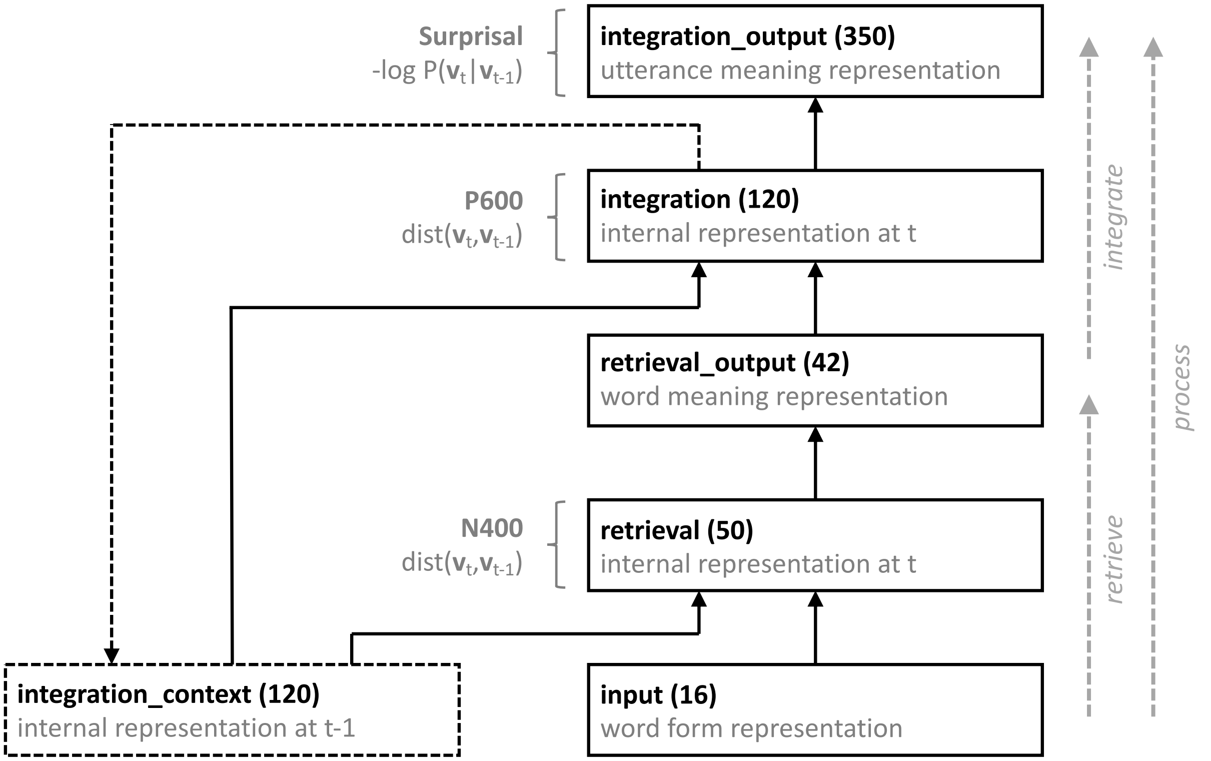

Surprisal
-log P($\mathbf{v}_t$|$\mathbf{v}_{t-1}$)
integration_output (350)
utterance meaning representation
P600
dist($\mathbf{v}_t$,$\mathbf{v}_{t-1}$)
integration (120)
internal representation at t
retrieval_output (42)
word meaning representation
N400
dist($\mathbf{v}_t$,$\mathbf{v}_{t-1}$)
retrieval (50)
internal representation at t
integration_context (120)
internal representation at t-1
input (16)
word form representation
integrate
retrieve
process

Figure 5: Schematic illustration of the neurocomputational model. Reproduced with permission (CC BY) from Brouwer et al. (2021).

The architecture of this model, depicted in Figure 5, is effectively an extended SRN: **input** ↦ **retrieval** ↦ **retrieval_output** ↦ **integration** ↦ **integration_output**, in which both the **retrieval** and **integration** layer receive additional input from an **integration_context** layer that contains the activation pattern of the **integration** layer at timestep *t-1*. The model processes sentences on a word-by-word basis, and mechanistically, the processing of a word w_t can be conceptualized as a function *process(word form, utterance context)* ↦ *utterance representation*, which maps an acoustic or orthographic *word form*, and the *utterance context* as established after processing words $w_1 \ldots w_{t\text{-}1}$, onto an *utterance representation*, an interpretation spanning word $w_1 \ldots w_t$.

N400. This mapping from *word form* onto an *utterance representation* does, however, involve an intermediate representation; that is, in line with the Retrieval-Integration account, it is assumed that a *word form* is first mapped onto *word meaning* while taking the *utterance context* into account: *retrieve(word form, utterance context)* ↦ *word meaning*. This *retrieve* function, which is assumed to underlie the N400 component of the ERP signal, is implemented by the first part of the SRN: **input** ↦ **retrieval** ↦ **retrieval_output**, which maps localist *word form* representations (**input**) onto distributed lexical-semantic *word meaning* representations (**retrieval_output**), while taking *utterance context* into account (**integration_context**). The **retrieval** layer effectuates this mapping, and N400 amplitude is estimated as the degree to which the activation pattern of this layer changes as the result of processing an incoming word w_t:

$$\text{(Eq 5) N400}(w_t) \; = \; dist(retrieval_t, retrieval_{t-1})$$

where $dist(x, y) \; = \; 1.0 \; - \; cos(x, y)$. Estimated N400 amplitude will be small if the model finds itself in a state in which the meaning of w_t is expected, as this will induce little change in the activation pattern of the **retrieval** layer from *t-1* to *t*. By contrast, if it is in a state in which the meaning of w_t is less expected, this will induce a larger change in the activation pattern of the **retrieval** layer, and consequently estimated N400 amplitude will be larger.

P600. Retrieved *word meaning* is subsequently integrated with the *utterance context* to produce an updated *utterance representation*, which can be conceptualized as the function *integrate(word meaning, utterance context)* ↦ *utterance representation*. This *integrate* function is hypothesized to underlie the P600 component and is implemented by the remainder of the SRN: **retrieval_output** ↦ **integration** ↦ **integration_output**, which maps the distributed lexical-semantic *word meaning* representation (**retrieval**), and the utterance context (**integration_context**), onto an updated utterance representation (**integration_output**). Here, the **integration** layer is responsible for this mapping, and P600 amplitude is therefore estimated as the degree to which the activation pattern at this layer changes as a result of processing a word w_t:

$$\text{(Eq 6) P600}(w_t) \;=\; dist(integration_t, integration_{t-1})$$

If the interpretation resulting from integrating word w_t is expected, given the input history of the model as well as the world knowledge contained within the meaning space, the activation pattern in the **integration** layer will change little from *t-1* to *t*, and estimated P600 will be small. If on the other hand, the resultant interpretation is less expected, a larger change in the activation pattern will ensue, and estimated P600 amplitude will be larger (also see Crocker et al., 2010).

Surprisal. Finally, like in the Venhuizen et al. (2019) model, surprisal is estimated as the negative log probability of the utterance meaning as constructed by the model after processing a word w_t -- that is, from the *utterance representation* produced at the **integration_output** layer -- given the utterance meaning as understood by the model prior to encountering w_t:

$$\text{(Eq 7) } surprisal(w_t) \;=\; -logP(integration\_output_t \mid integration\_output_{t-1})$$

Indeed, the model predicts a close link between the P600 and surprisal, where P600 amplitude indexes the work involved in updating an utterance meaning representation from *t-1* to *t*, and surprisal the likelihood of the resultant change in meaning.

To further evaluate the predictions of the model, Brouwer et al. (2021) modeled the N400 and P600 findings from a recent study by Delogu, Brouwer, and Crocker (2019), as well as the reading time data from a self-paced reading paradigm replication of this study in terms of surprisal. The design of this study differentially manipulated retrieval and integration difficulty through association and plausibility, respectively, while also avoiding the anomalous nature of the role-reversal evidence discussed above. When only plausibility is manipulated ("John [left/entered] the restaurant. Before long he opened the menu") – "menu" is semantically associated with "restaurant", but unlikely to be opened after having left versus entered a "restaurant" – a P600-effect ensues at "menu", reflecting increased integration difficulty, while the association results in no N400 effect being elicited. By contrast, when both association and plausibility are manipulated ("John entered the [apartment/restaurant]. Before long he opened the menu") – "menu" is both unassociated with "apartment" and unlikely to be opened in an "apartment" – an N400-effect is produced, reflecting increased retrieval difficulty, and this is followed by an occipitally-distributed P600-effect, reflecting increased integration difficulty.[3] In a simulation of this experiment, the model was shown to predict the same pattern of estimated ERP effects. Behaviorally, in turn, all contrasts increased RT at the target word (as well as in the

[3] This P600 effect is both stronger, and centro-parietally distributed, after correcting for spatiotemporal component overlap that arises due to a large N400 effect masking a subsequent P600 effect (Brouwer et al., 2021). In a follow-up study, Delogu et al., (2021) further confirm the presence of such a centro-parietal P600-effect for integration difficulty, when there is no confounding N400 difference preceding the predicted P600 effect.

spillover region). Following Hale (2001) and Levy (2008), the surprisal estimates of the model are taken to correspond to RTs. Crucially, the model also predicts the pattern of increased RTs for all contrasts. In sum, the neurocomputational model correctly predicted the N400-effects, P600-effects, and RT results, and confirmed the predicted relationship between the P600 and surprisal.

The implications of these modeling results extend well beyond this one particular study. First, the neurocomputational model offers a general, integrated algorithmic level account with explicit linking hypotheses to the N400, the P600, and surprisal/RTs in language comprehension, that is sensitive to probabilities manifest in both linguistic experience and knowledge about the world. Several neurocomputational models have focused on modeling the lexical retrieval processes underlying the N400 component alone based on evidence from the processing of words in isolation (e.g., Laszlo & Plaut, 2012; Rabovsky & McRae, 2014). These models provide important mechanistic explanations for a wide spectrum of lexical properties known to influence the N400, such as frequency, orthographic neighborhood size, and semantic relatedness, which offer furthers support for the view that the N400 is indeed an index of lexical retrieval processes. These mechanistic accounts are similar in nature and fully consistent with the instantiation of retrieval in the neurocomputational instantiation of the Retrieval-Integration model, which in its current form is focused on the additional contribution of sentence-level expectancy to retrieval processes.

Two more recent models have focused on modeling sentence level comprehension. In contrast to recovering a rich meaning representation, however, one model uses next word prediction as a proxy for comprehension (Fitz & Chang, 2019), rendering a direct comparison to proper comprehension models difficult. The other model is a comprehension model in which the

N400 is an index of the “quasi-compositional” mapping of sentences onto “sentence gestalt” representations (Rabovsky et al., 2018; Rabovsky & McClelland, 2019). Crucially, this quasi-compositional mapping is effectively “good enough” semantic integration, and the N400 amplitude induced by a word is a function of the degree to which the updated “sentence gestalt” is expected. On this model, the absence of an N400-effect for the plausibility-only manipulation (“John [left/entered] the restaurant. Before long he opened the menu”) in the Delogu et al. data would be accounted for by (temporarily) misunderstanding “left” as “entered” in the sentence prior to the target sentence, presumably because of the strong semantic association/attraction between “menu” and “restaurant”. While not explicitly part of their computational model, Rabovsky and McClelland (2019) suggest the P600 is a “more-controlled attention-dependent process” that subsequently resolves this temporary misunderstanding, correctly predicting the P600-effect for “John [left/entered] the restaurant. Before long he opened the menu”. Importantly, however, their model predicts only an N400-effect for “John entered the [apartment/restaurant]. Before long he opened the menu”, which is problematic as a P600-effect is also present for the latter contrast as discussed above. Finally, the contrast “John [entered/left] the restaurant. Before long he opened the umbrella”, produces a P600-effect and no N400-effect (Delogu et al., 2021). As there is no semantic association/attraction between “restaurant” and “umbrella”, it is unclear why “entered” should be misunderstood as “left” when reading the first clause, and hence how this result can be reconciled with the Rabovsky and McClelland account .

Given that the meaning representations that the Retrieval-Integration model recovers during comprehension derive from propositional co-occurrence, the coverage of the model can be scaled far beyond what is possible with simple, slot-based thematic-role assignment representations (cf. Crocker et al., 2010, Brouwer et al., 2017, Rabovsky et al., 2018, but see

Rabovsky et al., 2021, for an approach to scaling the sentence-gestalt approach). Hence, the model not only has broad coverage of neural and behavioral processing indices, but also in terms of the processing phenomena that it can capture; that is, the model is capable of capturing N400, P600, and surprisal/RT modulations driven by syntactic, semantic, and pragmatic aspects of incremental, word-by-word comprehension.

Secondly, the model also bridges the gap to functional-neuroanatomic models of language processing; that is, Brouwer et al. (2017) show how their neurocomputational model -- and thereby the model discussed above -- aligns with a minimal cortical processing network instantiating Retrieval-Integration cycles. This cortical network, depicted in Figure 6, is centered around two cortical *epicenters* or *hubs* --- the left posterior Middle Temporal Gyrus (lpMTG; Brodmann Area; BA 21) and left Inferior Frontal Gyrus (lIFG; BA 44/45/47) --- which are assumed to be core nodes in larger networks, serving as critical gateways for the integration of information from various sub-networks (see Brennan et al., 2020 for recent modeling evidence consistent with such a central role for these areas). More specifically, the lpMTG is taken as an epicenter/hub for Retrieval and is therefore the core generator of the N400 component. Integration, in turn, is subserved by the lIFG, and activity in this area is the presumed core generator of the P600 component. The lpMTG and the lIFG are wired together through white matter tracts in the dorsal pathway (dp) and the ventral pathway (vp). Figure 6 shows the alignment of the neurocomputational model to this cortical network. Depending on the input modality, incoming words enter the system through either the auditory cortex (ac) or the visual cortex (vc), corresponding to the **input** layer in the model. The lpMTG then serves to retrieve the meaning of an incoming word, while taking the unfolding context into account (lIFG ↦ lpMTG via either dp or vp), thereby generating the N400. The lpMTG aligns with the **retrieval** and

retrieval_output layers of the model, of which the former generates an N400 estimate, and receives the unfolding context through the recurrent projection from the **integration** layer.[4] Retrieved word meaning is then projected to the lIFG (lpMTG ↦ lIFG via either dp or vp), where it is integrated into the unfolding utterance interpretation, thereby generating the P600. The lIFG, in turn, aligns with **integration** and **integration_output** layers, of which the former generates a P600 estimate.

In sum, the neurocomputational model outlined above, 1) offers an integrated account of the N400, P600, and surprisal/RTs in incremental, word-by-word comprehension, 2) has the potential to scale up to a wide range of syntactic, semantic, and pragmatic processing phenomena, and 3) connects computational models of comprehension to functional-neuroanatomy, thereby paving way for an even more integrated investigation of language in the brain.

[4] Note that a shorthand notation is used for recurrent projections from **integration ↦ integration** and **integration ↦ retrieval**, and omit the **integration_context** layer from the figure.

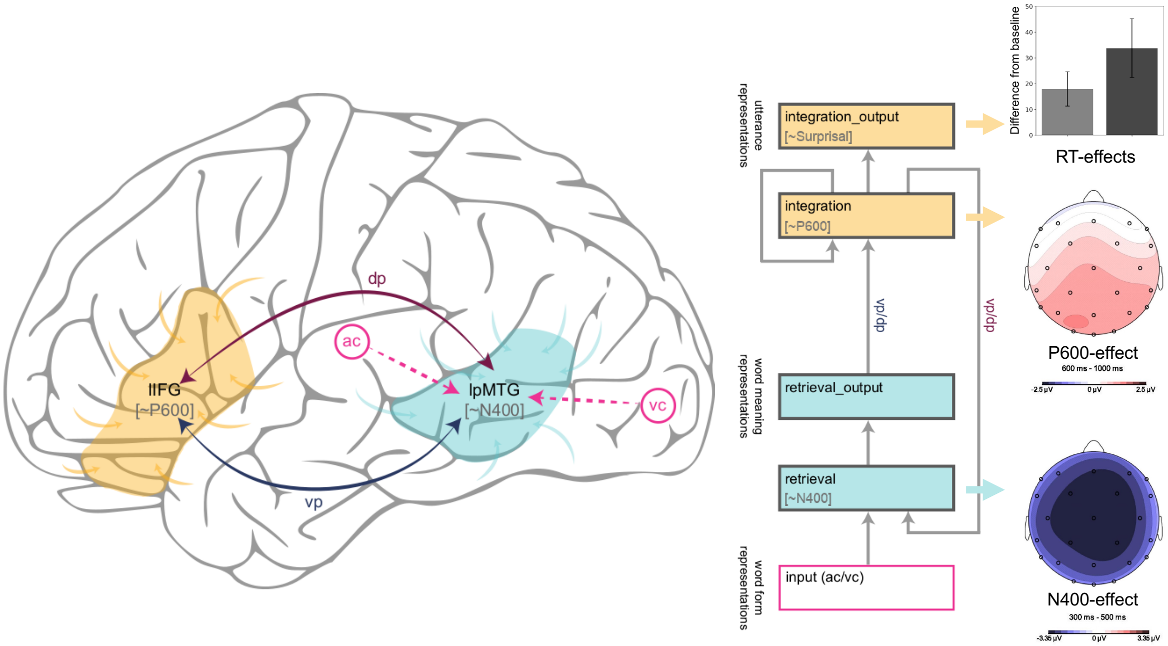


Figure 6: Alignment of the Neurocomputational Model to a minimal cortical network. Reproduced with permission from Brouwer et al. (2017, CC BY-NC), Brouwer et al. (2021, CC-BY), and Delogu et al. (2019, CC BY-NC-ND).

6. Conclusion

Models of human language comprehension seek to explain how people map the linguistic signal, word-by-word, into a representation of the intended meaning. Despite the complexity and ambiguity inherent in this task, it is something people mostly do effortlessly. While early theories were shaped by those situations in which people have difficulty – positing architectures and strategies aimed to limit demands on working memory or limit the influence of diverse information sources – there is increasing agreement that the language comprehension system can be viewed as rational. That is, in general, people rapidly deploy their prior experience with language and their knowledge of the world to probabilistically distribute their attention across possible interpretations of the unfolding utterance.

Thus, while many computational models have been proposed, the last twenty years have witnessed increasing consensus – at least at Marr's computational level – that the comprehension system seeks to maximize the likelihood of recovering the correct interpretation. This in turn has led to a linking hypothesis, Surprisal Theory, in which expected words are easier to process than unexpected ones. The range of phenomena that can be accounted for by surprisal is considerable, but this success is somewhat mitigated by the causal bottleneck – many different probabilistic mechanisms can yield accurate conditional word probabilities, including many that make no attempt to comprehend language, such as language models trained to simply predict the next word. Indeed, due to their simplicity and ease of training, such models often can provide a superior fit to empirical measures, but nonetheless say little about the actual comprehension mechanism that yields those measures. This emphasizes the point that modeling empirical measures must be secondary to modeling the task in question, namely language comprehension.

To this end a recent model was presented in more detail, which (a) utilizes rich probabilistic meaning representations that go beyond conventional syntactic parsing models, and further incorporate the influence of world knowledge, (b) implements an expectation-driven model in which surprisal is viewed as being "meaning-centric" measure of how difficult it is to integrate the current word into the unfolding representation of the utterance, and (c) provides transparent, mechanistic linking hypotheses to three distinct dependent measures that differentially index lexical retrieval (N400), semantic integration (P600), and overall cognitive effort (reading times) – each in a manner that is consistent with the expectation-driven nature of the system as whole, instantiating surprisal theory. More generally, this serves to illustrate how progress in cognitive modeling of language can benefit from combining rational theorizing about what the system computes and what kinds of representations are needed, with explicit links to multiple behavioral and neurophysiological empirical measures that differentially index the processes that recover those representations. Only by bringing to bear this combination of rational and empirical approaches to constrain and inform computational models and theories will it be possible to converge on closer approximations of the human language comprehension system.

Acknowledgments

The authors gratefully acknowledge the support of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 232722074 – SFB 1102. We also thank Christoph Aurnhammer, Francesca Delogu, Noortje Venhuizen, and four anonymous reviewers for their comments.

References

Alishahi, A. (2010). *Computational Modeling of Human Language Acquisition*, Morgan & Claypool.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98*(3), 409–429. https://doi.org/10.1037/0033-295X.98.3.409

Aurnhammer, C. & Frank, S. L. (2019). Comparing gated and simple recurrent neural network architectures as models of human sentence processing. *Proceedings of the 41st Annual Conference of the Cognitive Science Society*, Cognitive Science Society, Austin, 112-118

Baggio, G., & Hagoort, P. (2011). The balance between memory and unification in semantics: A dynamic account of the N400. *Language and Cognitive Processes, 26*, 1338–1367.

Bever, T. G. (1970). *The cognitive basis for linguistic structures. Cognition and the development of language.* ed. by John R. Hayes, 279–352. New York: Wiley.

Boston, M. F., Hale, J., Kliegl, R., Patil, U., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research, 2*, 1–12.

Bowman, S. R., Rastogi, A., Gupta, R., Manning, C. D., & Potts, C. (2016). A Fast Unified Model for Parsing and Sentence Understanding. *Proceedings of the Association for Computational Linguistics*, pp. 1466–1477.

Brennan J.R. & Hale J.T. (2019). Hierarchical structure guides rapid linguistic predictions during naturalistic listening. *PLoS ONE, 14*(1): e0207741. https://doi.org/10.1371/journal.pone.0207741

Brennan, J. R., Kuncoro, A., Dyer, C., & Hale, J. T. (2020) Localizing syntactic predictions using recurrent neural network grammars. *Neuropsychologia 146*, pp. 1074–1079 doi: 10.1016/j.neuropsychologia.2020.107479

Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. C. J. (2017). A Neurocomputational Model of the N400 and the P600 in Language Processing. *Cognitive Science, 41*(S6), pp. 1318–1352. doi: 10.1111/cogs.12461

Brouwer, H., Delogu, F, Venhuizen, N. J., & Crocker, M. W. (2021). Neurobehavioral Correlates of Surprisal in Language Comprehension: A Neurocomputational Model. *Frontiers in Psychology 12:110*. doi: 10.3389/ fpsyg.2021.615538

Chater, N., Crocker, M. W. & Pickering, M. J. (1998), The rational analysis of inquiry: The case for parsing, In: Nick Chater & M. Oaksford (eds.), *Rational Analysis of Cognition*, Oxford University Press, Oxford, (441–468).

Crocker, M. W. (1996). *Computational Psycholinguistics: An Interdisciplinary Approach to the Study of Language*, Dordrecht, NL: Kluwer.

Crocker, M. W. (1999), Mechanisms for sentence processing. In: Simon Garrod & Martin J. Pickering (eds.), *Language Processing*, Psychology Press, London, (191–232).

Crocker, M. W. & Brants, T. (2000). Wide Coverage Probabilistic Sentence Processing. *Journal of Psycholinguistic Research; 29*(6):647-669.

Crocker, M.W. (2005). Rational models of comprehension: Addressing the performance paradox. In A. Cutler (Ed.), *Twenty-First Century Psycholinguistics: Four Cornerstones*, pp. 363–380. Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Dell, G. S., & Cholin, J. (2012). Language production: Computational models. In M. J. Spivey, K. McRae, & M. F. Joanisse (Eds.), *The Cambridge handbook of psycholinguistics* (pp. 426–442). Cambridge University Press.

Delogu, F., Crocker, M. W. & Drenhaus, H. (2017). Teasing apart coercion and surprisal: Evidence from ERPs and eye-movements. *Cognition*, 161, 46-59.

Delogu, F., Brouwer, H., & Crocker, M. W. (2019). Event-related potentials index lexical retrieval (N400) and integration (P600) during language comprehension. *Brain and Cognition, 135*. doi: 10.1016/j.bandc.2019.05.007

Delogu, F., Brouwer, H., & Crocker, M. W. (2021). When components collide: Spatiotemporal overlap of the N400 and P600 in language comprehension. *Brain Research, 1766*. doi: 10.1016/j.brainres.2021.147514

Demberg, V. & Keller, F. (2008). Data from Eye-tracking Corpora as Evidence for Theories of Syntactic Processing Complexity. *Cognition* 109:2, 193-210.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14(2), 179–211. doi:10.1207/s15516709cog1402_1

Ferreira, F., V. Ferraro, and K. G. D. Bailey. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11, pp. 11–15.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology,* 47, pp. 164–203.

Ferreira, F. & Patson, N. (2007). The 'Good Enough' Approach to Language Comprehension. *Language and Linguistics Compass,* 1/1–2, pp. 71–83.

Fitz, H., & Chang, F. (2019). Language ERPs reflect learning through prediction error propagation. *Cognitive Psychology*,111,15–52. doi: 10.1016/j.cogpsych.2019.03.002

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, Mass: MIT Press.

Frank, S. L., Haselager, W. F., & van Rooij, I. (2009). Connectionist semantic systematicity. *Cognition, 110*(3), 358–379. doi:10.1016/j.cognition.2008.11.013

Frank, S. L., Koppen, M., Noordman, L. G., & Vonk, W. (2003). Modeling knowledge-based inferences in story comprehension. *Cognitive Science, 27*(6), 875–910. doi:10.1207/s15516709cog2706_3

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, *140*, 1-11.

Frazier, L. (1979). *On Comprehending Sentences: Syntactic Parsing Strategies*. Ph.D. thesis, University of Connecticut, Connecticut.

Gibson, E. A. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1–76.

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, *110*(20), 8051-8056.

Gibson E, Tan C, Futrell R, Mahowald K, Konieczny L, Hemforth B, Fedorenko E. (2017). Don't Underestimate the Benefits of Being Misunderstood. *Psychological Science*. 28(6):703-712. doi: 10.1177/0956797617690277

Gouvea, A.C., Phillips, C., Kazanina, N., Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes, 25*, 149–188.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In: *Proceedings of North American Association for Computational Linguistics*, Vol. 2, 159–166.

Hoeks, J. C. J., Stowe, L. A., and Doedens, G. (2004). Seeing words in context: The interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, 19(1):59–73.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science 20*: 137–94.

Kim, A. & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language, 52*(2):205–225.

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*, pp. 621-47.

Kutas, M. & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(4427):203–205.

Laszlo, S., & Plaut, D. C. (2012). A neurally plausible Parallel Distributed Processing model of event-related potential word reading data. *Brain and Language, 120*, 271–281. https://doi.org/10.1016/j.bandl.2011.09.001.

Lenci, A. (2018). Distributional Models of Word Meaning. *Annual Review of Linguistics*, 4(1):151-171

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

Lewis, R.L. & Vasishth, S. (2005). An Activation-Based Model of Sentence Processing as Skilled Memory Retrieval. *Cognitive Science*, 29, 375-419. https://doi.org/10.1207/s15516709cog0000_25

Linzen, T. & Baroni, M. (2021). Syntactic structure from deep learning. *Annual Reviews of Linguistics*. 7:195–212

Lopopolo, A. & Rabovsky, M. (2021). Predicting the N400 ERP component using the Sentence Gestalt model trained on a large scale corpus. *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society*.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*(4), 676–703. https://doi.org/10.1037/0033-295X.101.4.676

Magnuson, J. S., Mirman, D., & Harris, H. D. (2012). Computational models of spoken word recognition. In M. Spivey, K. McRae, & M. Joanisse (Eds.), *The Cambridge Handbook of Psycholinguistics*, pp. 76-103. Cambridge University Press.

Marcus, M. P. (1980). *A Theory of Syntactic Recognition for Natural Language*. Cambridge, MA: MIT Press.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W.H. Freeman.

Mayberry, M. R., Crocker, M. W., & Knoeferle, P. (2009). Learning to attend: A connectionist model of situated language comprehension. *Cognitive Science, 33*(3):449–496.

McClelland, J. L., St. John, M. F. & Taraban, R. (1989). Sentence comprehension: a parallel distributed processing approach. *Language and Cognitive Processes, 4*, pp. 287–336.

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*(3), pp. 283–312.

Michaelov, J. & Bergen, B. (2020). How well does surprisal explain N400 amplitude under different experimental conditions? In *Proceedings of the 24th Conference on Computational Natural Language Learning*.

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In: Chase, W. G. (Ed.). *Visual Information Processing: Proceedings of the Eighth Annual Carnegie Symposium on Cognition*, Held at the Carnegie-Mellon University, Pittsburgh, Pennsylvania, May 19, 1972. Academic Press.

Pado, Ulrike, Matthew W. Crocker, & Frank Keller (2009), A probabilistic model of semantic plausibility in sentence processing, *Cognitive Science, 33*, pp. 794–838.

Pereira, F. C. N. (1985). A new characterization of attachment preferences. In D. Dowty, L. Karttunen, & A. Zwicky (Eds), *Natural language parsing: Psychological, computational, and theoretical perspectives*. Cambridge: Cambridge University Press.

Pritchett, B. L. (1988). Garden Path Phenomena and the Grammatical Basis of Language Processing. *Language, 64,* pp. 539-576.

Rabovsky, M., & McRae, K. (2014). Simulating the N400 ERP component as semantic network error: Insights from a feature-based connectionist attractor model of word meaning. *Cognition, 132*, pp. 68–89. https://doi.org/10.1016/j.cognition.2014.03.010.

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

Rabovsky, M., & McClelland, J. L. (2019). Quasi-compositional mapping from form to meaning: a neural network-based approach to capturing neural responses during human language comprehension. *Philosophical Transactions of the Royal Society B: Biological Sciences, 375*(1791). doi: 10.1098/rstb.2019.0313

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing. *Journal of Verbal Learning and Verbal Behavior, 22*, pp. 358–374.

Rayner, K., & Well, A.D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, 3, pp. 504–509.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422.

Roark, B., Bachrach, A., Cardenas, C., & Pallier, C. (2009). Deriving lexical and syntactic expectation-based measures for psycholinguistic modeling via incremental top-down parsing. In *Proceedings of the 2009 conference on empirical methods in natural language processing*, 324–333.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536.

Sanford, A. J., Leuthold, H., Bohan, J., & Sanford, A. J. S. (2011) Anomalies at the Borderline of Awareness: An ERP Study. *Journal of Cognitive Neuroscience, 23*(3), pp. 514–523.

Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences, 6*(9), 382–386.

Shannon, C.E. (1948) A Mathematical Theory of Communication. *Bell System Technical Journal, 27*(3), pp. 379-423.

Smith, N. J. & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302–319.

Spivey, M., McRae, K., & Joanisse, M. (Eds.). (2012). *The Cambridge Handbook of Psycholinguistics* (Cambridge Handbooks in Psychology). Cambridge: Cambridge University Press.

Staudte, M., Ankener, C., Drenhaus, H., & Crocker, M. W. (2021). Graded expectations in visually situated comprehension: Costs and benefits as indexed by the N400. *Psychonomic Bulletin & Review*, 28, pp. 624–631.

Stevenson, S. (1994). Competition and recency in a hybrid network model of syntactic disambiguation. *Journal of Psycholinguistic Research, 23*(4), pp. 295-322.

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

Tanenhaus, M. K., Trueswell, J. C. & Hanna, J. E. (2000), Modeling thematic and discourse context effects with a multiple constraints approach: implications for the architecture of the language comprehension system. In: M. W. Crocker, Martin J. Pickering, & C. Clifton (eds.), *Architectures and mechanism for language processing*, Cambridge University Press, Cambridge, pp. 90–118.

Taylor, W. L. (1953). "Cloze procedure": a new tool for measuring readability. *Journalism Quarterly, 30*, pp. 415–433.

Townsend, D., & Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules*. Cambridge, MA: MIT Press.

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*, pp. 285– 318.

van Dijk, T. A., & Kintsch, W. (1983). Strategies of discourse comprehension. New York: Academic Press.

van Herten, M., Kolk, H. H. J., and Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, 22(2):241–255.

Venhuizen, N. J., Crocker, M. W., & Brouwer, H. (2019). Expectation-based Comprehension: Modeling the interaction of world knowledge and linguistic experience. *Discourse Processes, 56*:3, pp. 229–255. doi: 10.1080/ 0163853X.2018.1448677

Venhuizen, N. J., Hendriks, P, Crocker, M. W., & Brouwer, H. (2021). Distributional Formal Semantics. *Information and Computation*. https://doi.org/10.1016/j.ic.2021.104763

Warren, T. and Dickey, M.W. (2021), The use of linguistic and world knowledge in language processing. *Language and Linguistics Compass*, 15: e12411. https://doi.org/10.1111/lnc3.12411

Wehbe L, Murphy B, Talukdar P, Fyshe A, Ramdas A, Mitchell T (2014). Simultaneously Uncovering the Patterns of Brain Regions Involved in Different Story Reading Subprocesses. *PLoS ONE* 9(11): e112575.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*(2), 162–185. doi:10.1037/0033-2909.123.2.162